Understanding the Role of Packing Defects through Combinatorial Engineering of a Four-Helix Bundle Protein

Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Bachelors of Science with Honors Research Distinction in the College of Engineering of The Ohio State University

By

Jaron Larry Lohmeyer, B.S.
Undergraduate Program in Biomedical Engineering

The Ohio State University

2016

Thesis Defense Committee:
Dr. Thomas J. Magliery, Advisor
Dr. Gunjan Agarwal, Advisor
Dr. Jennifer Leight

Abstract

Library characterization is a useful method to understand the structural principles that govern tertiary and quaternary folding of proteins. Model proteins, such as Rop, are frequently used to facilitate better insight to these structural motifs. Rop is a small sixty-three residue protein that forms an antiparallel, homodimeric four helix bundle. It regulates plasmid copy number by facilitating binding of RNA I to the priming RNA II of ColE1 origin plasmids, which makes it favorable for *in vivo* cell based activity screens. The hydrophobic core consists of an eight layer heptad repeat pattern and studies have demonstrated that perturbation of the core can drastically alter RNA binding ability, shift the thermal stability, and change the antiparallel/parallel conformation. The packing pattern at the Arg-55 and Phe-56 positions is reversed from the regular heptad repeat and it is not well understood as to why the pattern shifts. Some studies have suggested that the phenylalanine acts as a hydrophobic plug to the core, but do not fully explain the structural impacts. This investigation, based on preliminary results from the R55 point study, made a combinatorial library and engineered non-native variants to better understand the packing effects of the 55^{th} and 56^{th} positions. The variants were classified by activity and characterized through thermal and chemical melts. The correlation of stability data with structural information from crystallography and NMR further defined the role of Arg-55 and Phe-56 and quantify core packing constraints.

Dedication

Thank you to my parents, Steve and Susan, brother, Cory, and girlfriend, Brittany, for all of your love and support.

Acknowledgements:

Dr. Thomas J. Magliery:
Dr. Gunjan Agarwal:
Dr. Jennifer Leight
Graduate Student Anusha Kumar
Graduate Student Nishanthi Panneerselvam

Funding:
Dr. Thomas J. Magliery
Ohio State Undergraduate Education Summer Research Fellowship
Ohio State College of Engineering Undergraduate Research Scholarship
NIH R01 Grant

Vita

May 2011……Louisville High School

2011 to present……Student, Department of Biomedical Engineering, The Ohio State University

Field of Study

Major Field: Biomedical Engineering

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Protein Engineering

What are the driving forces that make a protein fold into its destined conformation? How do small perturbations in a local structure influence the structure and function of the whole protein? Are these minor mutations really that important? Proteins serve many vital structural and functional roles throughout the body. Anfinsen determined through his work with RNase that the primary structure of the protein determines its final tertiary structure.[1] It seems very simplistic, but considering there are 20 natural amino acids, even a small protein with 50 residues has approximately 50^{20} possible combinations. Then within the number of amino acid combinations there are nearly infinite number of possible interactions. The vastness of protein folding makes it very difficult to accurately predict the structure and function of a primary sequence. There have been great strides in recent years, utilizing computational methods in combination with experimental methods, that have refined our understanding of protein folding but we are still unable to accurately predict the final destination of a primary sequence.[2] There have been successful predictions of structural and functional alterations due to mutagenesis and researchers such as David Baker, at the University of Washington, have been able to successful predict the structure of compatible binding sites with high affinity.[3]

The ability to create *de novo* proteins to restore function or fulfill a novel role is ultimately the goal of protein engineering. Developing novel proteins that have native-like structure and function, but provide higher thermal and chemical tolerances have many biomedical applications. Extended drug storage, different drug delivery pathways, and treatment of pathologies with protein based therapies are some of the examples. Many detrimental diseases are caused by genetic mutations that ultimately translate to non-functional or partially functioning proteins. Even small mutations can have detrimental impacts to the structure and function of a protein. Cystic Fibrosis (CF) is an example of this, which can be caused by a single mutation, an F508 deletion in the cystic fibrosis transmembrane conductance regulator (CFTR).[4]

1.2 Protein Folding

Protein engineering has two goals, 1) understand how proteins fold into their native state and 2) understand the driving forces that dictate native conformational stability and function. Folding pathways provide insight into how proteins can test an astronomically large number of conformations and reach a native state almost instantly. Protein folding is thought to happen according to a folding funnel pathway. The funnel is a basic representation of the number of conformations that the protein can take on as dictated by amount of free change. The top of the funnel represents a high energy state and as a result the protein has many conformational states. The funnel narrows as the amount of free energy decreases because the low energy states of the protein have fewer conformations. The change in free energy along the protein folding pathway is a result of changes in intermolecular and atomic forces within the protein and between the protein and solvent. The funnel is not smooth however and it can have local minima surrounded by local maxima. The local minima create a partially folded protein known as an intermediate.[5,6] The law of thermodynamics, achieving the lowest free energy state, pushes the protein folding down the funnel until a minimum is reached. The final state minimum contains the natively folded protein, but the final state minimum is not necessarily the lowest energy state within the entire energy landscape. It could instead be the lowest energy state achievable by the protein because of the energetic barriers around the minimum on the folding pathway. The energetic difference between natively folded proteins and molten globules, partially denatured structures, can be as small as 5 – 10 kcal/mol.[7]

The generally accepted protein folding pathway is initiated by hydrophobic interactions, which causes what some refer to as 'hydrophobic collapse'. From this conformational state local secondary structure begin to form, which then interact with other local secondary structures in a local-to-global[8] chain of events that then forms the tertiary structure of the protein. From this theory of protein folding, hydrophobic interactions are extremely important in determining the final native conformation. The core of a protein is formed by nonpolar amino acids that are sequestered from water and favorably stacked together to form favorable van der Waals interactions. Theories of protein core packing, oil-drop and jigsaw conundrum, have argued over the importance of

hydrophobic sequestration versus optimizing the formation of van der Waals interactions, but both theories agree on the importance of protein core interactions.[9,10] Hydrogen bonding is also very important in protein structure and function. Secondary structures are formed through hydrogen bonding of the peptide backbone and the external portions of the protein typically optimize hydrogen bonding with the solvent. Backbone angle preferences play a large role in protein conformation and secondary structure formation. Backbone angle preferences are defined as Phi and Psi angles and favorable angles for amino acids are determined by steric and rotational energetic factors. Electrostatic interactions affect protein structure and function because of the attractive and repulsive forces generated by positive and negative charges. Chain entropy opposes protein folding because a natively folded protein has a restricted conformation and an ordered state.

1.3 Combinatorial Studies

1.3a Approaches to Protein Engineering

Anfinsen conducted the first protein folding study on RNase A by using different concentrations of a chemical denaturant. His discovery and methods were revolutionary for understanding protein folding pathways and the stabilizing interactions of protein folding and function. However, the methods were limited and quantifying the interactions between structure and function, especially for site specific interactions was near impossible. In 1978, the first site-directed mutagenesis was conducted and soon after applied to functional enzymes to study the interactions and importance of active site residues.[11-13] The ability to conduct site-directed mutagenesis in combination with solving crystallization structures essentially created the field of protein engineering by enabling the study of governing interactions that accounted for the function, stability, folding, specificity, etc.[14,15] The introduction of site-directed mutagenesis was the 'tip of the iceburg' for the protein engineering community and new methods were developed to study amino acid interactions through combinatorial studies, cassette mutagenesis, or directed evolution for randomly mutated genes.[16] The more modern techniques were more compatible with high throughput methods, which enabled more thorough investigations governing protein interactions. Modern approaches to protein engineering also involve computational modeling of proteins. Computational model predict protein

structure by estimating forcefield energies and iteratively optimizing 3D conformations that correspond with low free energy states.[8] Computational models have become accurate enough to predict protein structure to guide further biochemical studies or aid with drug-discovery programs.[8] However, computational models are still very limited within the scope of protein engineering and rely heavily on previously solved structures in the protein data bank (PDB).[8] New technologies are constantly being developed for computational modeling, such as the fragment assembly approach, but it is still essential to understand and quantify the specific interactions that govern protein folding, structure, and function.[8]

1.3b Screening Methods

The protein engineering community initially took on the protein folding problem through systematic mutation of amino acids in sequence. The technique is generally applied to well-studied proteins that can be characterized through biophysical analyses. A common method to investigate the surface, core, loop, and active site interactions in a protein is a combinatorial approach that randomizes the nucleotide sequence within desired codons. The combinatorial approach is much higher throughput than the site-directed mutagenesis studies mentioned earlier. Site-directed mutagenesis studies do not require screening for active variants, but as a result are much slower to characterize and the native-like function cannot be confirmed. Combinatorial biophysical characterizations are effective because they generate large libraries of site-directed mutants that can be screened for stability, structure, and function.[17-19] It is an ideal method to collect large data sets on a protein because of its compatibility with high throughput analysis.

There are three primary elements of a combinatorial approach. First, there is a construction of library variants. The construction of libraries is easily achieved through modern day PCR techniques that utilize mixes of nucleotides to randomize desired codon sequences. The second step of a combinatorial approach is the selection or screening of the library molecules for the desired 'native-like' protein. Combinatorial selection and screening techniques attempt to tie genotype to phenotype and have historically achieved this through genetic screening or turn-over of chromogenic molecules.[17] Genetic selection is a screening technique that ties the protein's function to a vital cellular mechanism. Studies of tryptophan synthase have utilized tryptophan-free media to select

for functional protein, and the protein variants that did not have native-like function resulted in cell death. The problem with the genetic selection approach is that non-functioning protein cannot be studied because the function of the protein is linked to the survival of the cell.[17] In cases where the function of the protein cannot be tied to the survival of the cell, a different approach has to be taken. The basic function of a screen is to link the functionality of the protein to an observable phenotype. The turnover of chromogenic materials is commonly used because it allows for the classification of protein variants on both ends of the functional spectrum, from native-like to molten globule. The drawback of screening for variants in this fashion is that it requires a certain population of the protein variants to have native-like function.[17] After the protein has been screened or selected for, the selected variants must be identified. Proteins that are expressed from plasmid DNA can be sequenced through the assumption that the genetic code directly encodes the primary sequence of the protein. Another method for identifying the protein is through phage-display. Phage-display utilizes the display of proteins on the surface of a bacteriophage, which can be immobilized on a ligand and unfolded or non-binding protein can be washed away. The protein of interest can then be eluted separately and the DNA of the bacteriophage can be analyzed for the protein sequence.

1.3c Biophysical Characterization

Combinatorial library studies are able to generate and screen large amounts of variants. The screen of combinatorial studies allows scientists to relate structure to function. It is also important to understand how structure relates to the stability of the protein. Biophysical characterization techniques make it possible to evaluate stability aspects of mutated variants. The stability of a protein is the result of the free energy between the folded and unfolded state. Denaturation experiments observe the transition from the folded to unfolded state, from known input environments, which allows the stability of variants to be calculated. Proteins can be denatured through thermal input, heating the samples, or through chemical denaturation by concentrated urea, guanidine, changes in salt concentration, or changes in pH. The denaturation of a protein is typically monitored through circular dichroism (CD) or fluorescence. CD measures the absorbance of left and right polarized light across a spectrum of wavelengths. Chiral molecules

absorb polarized light and so do chiral protein secondary structures. Protein secondary structures, α-helicies and β-sheets, absorb polarized light differently. The denaturation of a protein causes secondary structure to collapse and the collapse can be monitored by the attenuation of polarized light absorption. Monitoring protein denaturation with fluorescence can use a specialized dye that fluoresces in low dielectric environments. As a protein is denatured the hydrophobic core becomes exposed to the solvent and the hydrophobic interaction between the dye and protein residues generates a fluorescence signal.

Combinatorial protein studies generate large libraries and require a high throughput biophysical characterization approach. A high throughput thermal scanning technique was developed by the Magliery Lab to evaluate the stability of proteins.[20] The principle of this technique utilizes the dye SYPRO orange, which fluoresces in low dielectric environments. Well folded variants of Rop produce low fluorescence signals in the presence of SYPRO orange. As the protein is heated SYPRO orange begins binding exposed hydrophobic residues and generates a fluorescence signal that correlates to the fraction of protein in the unfolded state. The thermal melts are conducted with a real-time PCR machine because most thermal denaturation experiments with Rop require temperatures of 90-95^{o}C. The real-time PCR machines are capable of holding a 96-well plate, which allows for 96 variants of Rop to be monitored at once. Small scale expression and purification procedures were developed for Rop variants. The technique was not used for characterization of the R55F56 library because the library only contained 400 possible variants.

The purpose of a Gibbs-Helmholtz analysis is to measure ΔG, the change in free energy between the folded state and unfolded state as a function of the temperature. The change in ΔG as a function of temperature indicates the conformational stability of a protein and the enthalpic stabilization. Since the unfolding process is defined as a two-step process, the ΔG, ΔH, and ΔS can be determined experimentally. The Gibbs free energy equation in standard form is $\Delta G = \Delta H - T\Delta S$ (1). By knowing that Gibbs free energy can also be represented by $\Delta G = -RT \ln K$ (2), the differential form of the van't Hoff equation $(\partial \ln K)/(\partial T) = \Delta H / (RT^2)$ can be substituted into to yield the differential form of the Gibbs-Helmholtz equation, $\partial(\Delta G/T)/\partial T = -\Delta H/T^2$. In biological reactions the

heat capacity of an interaction is not constant and can be calculated by taking the partial derivative of enthalpy with respect to temperature, $\partial\Delta H/\partial T = \Delta C_P$. From the partial derivative relationship with ΔC_P the following equations can be found: $\Delta H(T) = \Delta H_r + \Delta C_P(T-T_r)$ (2) and $\Delta S(T) = \Delta S_r + \Delta C_P \ln(T/T_r)$ (3). T_r is a reference temperature and can be used to solve the entropy and enthalpy at that reference temperature as well as ΔC_P. At the melting temperature, T_m, of a protein the $\Delta G = 0$, and this combined with equation (1) yields $\Delta S_m = \Delta H_m/T_m$ (4). Substituting equations (2), (3), and (4) into equation (1) gives a commonly used Gibbs-Helmholtz analysis $\Delta G = \Delta H_m [(T_m - T)/T_m] - \Delta C_p [T_m - T(1 - \ln(T_m/T))]$ (5).[21] The Gibbs-Helmholtz analysis used for the study of the R55 and R55F56 library variants was slightly different and used an analysis from Becktel and Schellman developed in 1987.[22]

1.4 Rop Structure and Function

Naturally, cells have a copy number regulation system that preserves their genetic make-up. Unnecessary replication of genetic information could introduce detrimental genetic mutations and requires a higher energetic demand on the cell. *E. coli* was initially discovered to have a regulation mechanism that controlled the copy number of plasmid DNA. RNA II was identified to be a 555 nucleotide structure that served as a primer for DNA polymerase I at the origin.[23] RNA I regulates the priming action of RNA II by forming a hybrid structure with RNA II through a 'zip mechanism'.[24] Rop was initially discovered in 1982 and identified as s 63 amino acid peptide that functioned as a regulator of ColE1 DNA replication.[25] It was found in 1984 that the function of Rop was to facilitate the zipping mechanism of the inhibitor RNA I to the priming RNA II, which further regulated the ColE1 plasmid copy number.[26-28]

The structure of Rop has been solved in crystals and in solution.[29,30] The structural solutions have shown that Rop is a 63 amino acid that forms an antiparallel homodimer. The structural motif is a coiled-coil that consists of a four-helix bundle, each monomer consists of two helices connected by a loop. The hydrophobic core is defined as an eight layer heptad repeat packing pattern and is shown below in Figure 1.1.[29] Each monomer's core layer is composed of one smaller hydrophobic amino acid dubbed position "a" and one larger hydrophobic amino acid dubbed position "d". When the monomers fold

together and form a dimer the small and large residues stack on top of one another, forming a well-structured tightly packed core.[31-33] The repetitiveness and simplicity of Rop's core has been the focus of many repacking studies. It has been debated as to how a protein achieves native stability, whether by dominance of hydrophobic surface burial or achieving favorable van der Waals packing. However, it is well understood that the composition of the hydrophobic core is one of the most important determinants of structure and stability. It is for this reason that *de novo* design and redesign of the core has been the focus of many structural studies. The design studies attempt to explain the structural and consequently functional impacts of residue substitutions and therefore model protein are typically utilized.

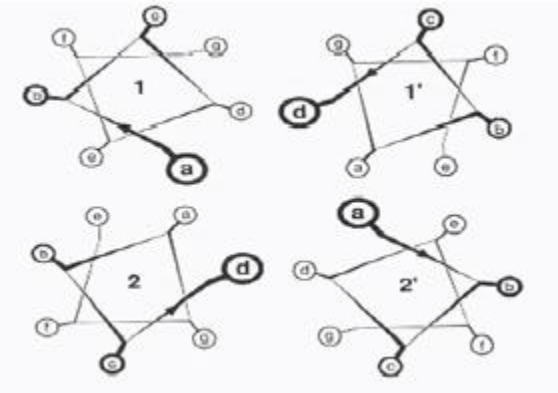


Figure 1.1: Heptad Repeat Packing Diagram. The diagram shows the four helixes that makeup Rop, with the "a" and "d" residues that favorably stack in the core. The 1 and 2 show one monomer and the 1' and 2' show the antiparallel second monomer.[33]

Model proteins are used because they rule out many of the complexities found in protein architecture that can complicate the perceived effects of mutated residues.[32,33] Rop is an excellent model for investigating core interactions because it is small, regular in structure, easily expressible in large quantities, has been linked to an *in* vivo functional assay, is compatible with high throughput studies, and the structure has been previously solved.[17,29,30] In Magliery Lab, Rop is expressed from a pMRH6TEV expression vector, which has a T7lac promoter and therefore can be induced with IPTG. The His tag enables for chelation of nickel agarose beads during purification, which can be hydrolyzed from the protein at the TEV site. An *in vivo* screen for Rop was developed by Magliery and Regan that linked ColE1 plasmid copy number to fluorescence through the regulated expression of GFP.[17]

1.5 Functional Assay for Rop

It is advantageous in combinatorial library examinations to have a screening method that allows for high throughput evaluation and classification of activity. To achieve this it is optimal that the protein be able to be evaluated *in vivo* without the need for purification. It is quite trivial why it is ideal not to purify the protein in order to analyze the activity when considering high-throughput methods. The purification process takes more time than what is required to evaluate the protein *in vivo*. More importantly, classifying the activity *in vivo* demands the functionality of the protein and structural similarities at the atomic level to 'native-like' proteins, which *in vitro* testing might not be able to demand. The screening technique that was utilized in this study was developed and reported by Magliery and Regan in 2004. The screening technique was a novel approach to evaluate Rop *in vivo* and directly relate the activity of the variants to a fluorescence signal. There were previous selection and screening methods reported for Rop, but those screens were not as robust or as high-throughput.[25,34-38]

The initial Rop screen was a negative screen developed by Cesareni. It utilized the fusion of β-galactosidase to the first 110 nucleotides of RNA II.[25,35] Functional Rop variants facilitated the binding of the inhibitory RNA I to the priming RNA II, which reduced the number of binding sites for the β-galactosidase.[26] A negative screen is not an optimal selection tool because there can be many causes for the lack of β-galactosidase function, such as a simple mutation in the β-galactosidase DNA sequence. A selection method for Rop was developed by Castagnoli through the fusion of Rop to the DNA binding domain of a λ repressor.[34] Rop variants that were able to dimerize caused immunity to the λ infection, but this also caused the cells with inactive Rop variants to die. An *in vitro* screen was developed that utilized RNA 'kiss complexes', isolated stem loops from RNA I and RNA II, which active Rop variants could bind and then be identified through an electrophoretic mobility shift assay.[36-38] It was later shown through the *in vivo* screen developed by Magliery and Regan that the electrophoretic mobility shift assay did not require as stringent functional and structural constraints as were required for Rop to be active *in vivo*.[17]

The *in vivo* screen developed by Magliery and Regan utilized the native function of Rop as a repressor of primer for plasmids with a ColE1 origin.[17] The screen is advantageous for many reasons, it can be done *in vivo*, the screen can be directly linked to the fluorescence from green fluorescent protein (GFP), and it has an easily distinguishable range of fluorescence between wild type Rop and nonfunctional variants. As was stated earlier, it is favorable to screen for variants *in vivo* because it is more compatible with high-throughput combinatorial library methods and in this case it directly relies on the protein's native function, which is then linked to a reporter molecule, GFP. The positive screen was created by the use of two separate plasmids. The Rop gene is contained on a pAClacRop plasmid, which also carries kanamycin resistance gene and has a p15A origin. The GFP expressing plasmid, pUCBADGFPuv, contained the GFP gene along with an arabinose promoter and an AraC gene.[17] The two screening plasmids can be seen below in Figure 1.2. The pUCBADGFPuv plasmid was created from pBAD-GFPuv, excising the pBR322 ColE1 origin and replacing it with a ColE1 origin from pUC19.[39] The pBAD-GFPuv plasmid is not innately a high enough copy plasmid to satisfy the requirements of the screen because it is a pBR322 derivative, which contains a cryptic promoter (GUG).[17] It was favorable to place the GFP gene in a high copy number plasmid because the copy number range was greater between wild type Rop and inactive Rop.

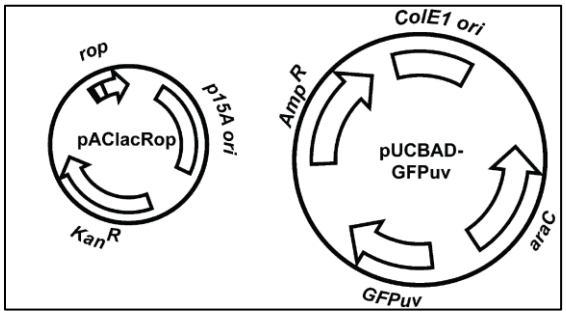


Figure 1.2: pAC and pUC Screening Plasmids. The pAC plasmid contains the Rop gene (left) and pUC GFP expression plasmid with ColE1 origin[17]

The cells with active Rop variants conferred a lower copy number of the pUCBADGFPuv plasmid than the cells with inactive variants. Intuitively, one would think that the cell with a high copy number would innately express more GFP. However, the AraC gene acts as a repressor to GFP expression in the absence of arabinose and a

promoter in the presence. When the cells were grown on solid state LB media that contained 0.0005% arabinose the cells with active Rop variants expressed GFP and fluoresced, as seen in Figure 1.3. It was hypothesized that the 0.005% arabinose was enough to induce promotion by the AraC gene when active variants regulated the copy number of the P_{BAD} plasmid, but the arabinose concentration was below the threshold for cells that contained inactive variants because the copy number was much higher. In addition to evaluating fluorescence, the amount of plasmid DNA was quantified through an alkyline lysis miniprep.

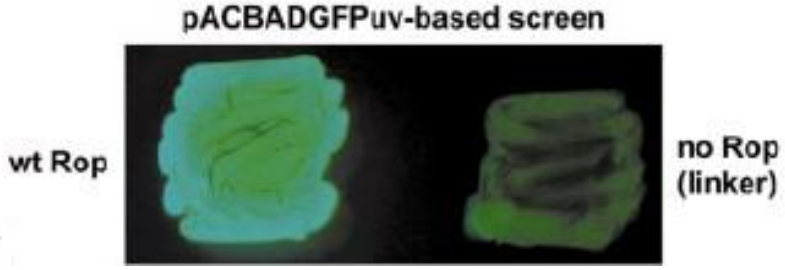


Figure 1.3: Positive screening results from the Magliery and Regan cell based screen[17]

1.6 Hydrophobic Interactions in the Core of Rop

The hydrophobic core is extremely important in determining the structure and function and understanding these interactions is essential for *de novo* design and redesign.[29,32,33,40,41] The only way to better understand the protein core forces that dictate protein architecture is by producing many core variants and then taking direct physical measurements.[18,42] Previous studies have found that repacking the core residue positions have the potential to substantially perturb the stability, function, topology, and thermally or chemically induced unfolding properties.[32,33,42] There are many energetic consequences, both enthalpic and entropic, that are associated with nonpolar amino acids. According to the laws of thermodynamics, the protein will adopt a conformation in the lowest free energy state that can be achieved through the specific folding pathway. The change in free energy, ΔG, is determined by the changes in enthalpy, ΔH, and entropy, ΔS, in the equation $\Delta G = \Delta H - T\Delta S$. When considering the energetic consequences of protein folding, evaluating ΔH and ΔS separately can help illustrate all of the forces at play.[43]

$$\Delta S = \Delta S^{conf} + (\Delta S_{pol} + \Delta S_{npl})^{hyd}$$

$$\Delta H = (\Delta H^{HB} + \Delta H^{hyd})_{pol} + (\Delta H^{vdW} + \Delta H^{hyd})_{npl}$$

The change in entropy is a result of the change in entropy due to protein fold conformation (ΔS^{conf}) plus the change in entropy due to solvation effects of both polar and nonpolar amino acids. The change in enthalpy is a result of the change in enthalpy of polar residues, considering hydrogen bonding and desolvation effects, plus the change in enthalpy of nonpolar residues, accounting for van der Waals interactions and desolvation effects.

There are multiple theories that suggest how the core of a protein forms. The oil drop theory states that the most important aspect of core packing is the burial of hydrophobic residues away from polar solvents.[9] According to this theory, nonpolar side chains might not achieve the tightest possible core volume or optimize van der Waals interactions. The purpose of core formation is to minimize unfavorable interactions between nonpolar and polar molecules. Slightly contrasting the Oil-droplet theory of core packing is the jigsaw puzzle approach to core packing.[10] The core theoretically packs by achieving favorable van der Waals interactions and tight side chain packing that minimizes the amount of voids within the core. The "knobs and holes" effect is one method in which a protein core can create and fill voids in an optimized fashion.[44]

The four-helix bundle protein Rop is a model protein to study the effects of core packing. The core of Rop consists of eight layers and the symmetric heptad repeat pattern, commonly found in coiled-coils, packs the "a" and "d" positions into the core.[31] As was mentioned above, wild-type Rop has smaller amino acids in the "a" position and larger amino acids in the "d" position, except for the second and seventh layer where the trend is switched. The reversal in the second and seventh layers is thought to cause an "end-effect", which results in a slight curvature at the ends of the bundles, and has been suggested to influence RNA binding.[33]

Previous studies that have repacked the core of Rop have substituted in amino acids with different propensities to forming α-helices, altered the core volume, or optimized packing sizes.[32,33,44] The studies repacked the core positions with Ala, Val, Leu, Ile, or Met and variants were made by repacking 2, 4, 6, or all 8 layers of the core. In addition, "-rev" variants reversed the size of the "a" and "d" positions in the 2nd and 7th layers to simulate wild-type Rop. The overpacked variant, Leu_4-8, repacked the entire

core with Leucine, which increased the core volume to 1,350 $Å^3$ per monomer from the 1,060 $Å^3$ of wild-type Rop.[33,45,46] The overpacked Rop core did not allow Rop to bind RNA hairpin loops in a gel mobility shift assay and analytical ultracentrifugation suggested that the variant formed a tetramer. However, the protein is extremely resistant to thermal and chemical denaturation and the increase in stability was thought to be caused by the increase in buried hydrophobic surface area. The underpacked variants, Ala_2Val_2 and Ala_4, were founded to be high destabilized, which suggested that the variants could not bury enough hydrophobic volume, the core volumes for Ala_2Val_2 and Ala_4 were, 940 $Å^3$ and 740 $Å^3$ per monomer, respectively.[47] The Ala_2Ile_2-6 variant was found to have a reoriented monomer so that the dimer was in a *syn* conformation and the loops were at the same end of the structure.[44] The structural solutions of the variants showed an offset core that allowed for a much tighter, staggered packing pattern "knobs-and-holes", than the parallel packing pattern found in wild-type.[44] The variant was inactive because the change in conformation split the RNA binding site, but the favorable core packing increased the thermal stability from that of wild-type Rop. The Ala_2Leu_2-2, 4, 6, and 8 and Ala_2Leu_2-8-rev variants conserved the buried hydrophobic volume and surface area of wild-type and follow the small "a" large "d" steric packing pattern, the Ala_2Leu_2-8-rev was the only variant that full emulated wild-type and switched the "a" and "d" position in the 2^{nd} and 7^{th} layers. The variants showed that systematically repacking the core with residues that simulated the properties of wild-type produced variants with high helical content, high resistance to thermal and chemical denaturation, and an ability to bind to RNA hairpin loops in a gel mobility shift assay.[33] However, an *in vivo* functional assay showed that even though the variants were able to bind the *in vitro* RNA hairpin loops, the variants were not able to functionally regulate the plasmid copy number.[17] The Ala_2Leu_2-8-rev and Ala_2Leu_2-8 variants, which were found to have a T_m much higher than that of wild-type Rop and efficiently bind RNA *in vitro*, were completely inactive *in vivo*.[17,33] The Ala_2Leu_2-4 variant was able to retain slight *in vivo* activity and Ala_2Leu_2-2, the most conserved variant, was able to retain strong *in vivo* activity. The ability to bind *in vitro* RNA, but lose *in vivo* activity showed that even small perturbations in the core of Rop can influence the atomic-level detail and native-like fold that was required for function.[17]

1.7 The R55F56 Conundrum

The core of Rop at the Arg-55 Phe-56 positions is interesting because the heptad repeat packing pattern, observed throughout the rest of the structure, is broken. The Arg-55 is a "d" position and should be found in the core, but structural studies have shown that Phe-56, a larger aromatic, packs the terminal layer of the core.[28] Site-directed mutation studies have attributed the packing to both enthalpic and entropic motivators, and hypothesized the position change is caused by the role of Phe-56 as a cork that fills the hydrophobic hole.[28,43] This structural difference in the heptad repeat can be seen in Figure 1.4. However, there have been no conclusive studies that have thoroughly investigated the influence of Arg-55 and Phe-56 on protein stability, structure, topology, and function.

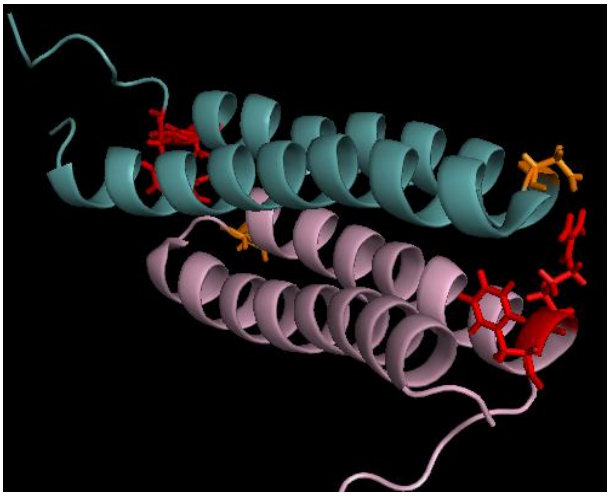


Figure 1.4: R55F56 Packing and Interaction with D32. Red residues highlight the R55F56 region, showing the R56 in the core of the protein and the F55 forming an electrostatic, stabilizing interaction with the D32 residue in orange. Figure was borrowed from PyMol with PDB fetch name 1RPR.

Evaluation of previously solved crystal structures suggests that the R55 residue interacts with D32, located in the loop of the adjacent monomer, which would create an electrostatic interaction and partially stabilize the terminal end of Rop. The R55 positive charge could potentially be interacting with two different residues in the loop D30 or D32, but the steric constraints observed in the solved structure of Rop suggest the primary electrostatic interaction is with D32. The interesting orientation of the R55F56 residue leads to some interesting questions. One would theorize that since the R55 and

D32 are in such close proximity to one another that the residues could potentially have destabilizing effects if R55 was mutated to a negatively charged residue. There are expected enthalpy shifts when mutating the R55 to non-basic or acidic residues because this would eliminate or repel the hypothesized electrostatic interaction between Asp-32 and Arg-55. Mutation of the 55^{th} position to a hydrophobic amino acid could potential to form favorable van der Waals interactions with the terminal end of the core. Also, the core of Rop has been subjected to repacking studies that have discovered interesting results when overpacking and underpacking the core. The entropic and enthalpic shifts caused by under-packing or over-packing the 55^{th} and 56^{th} positions should provide a clearer picture on the steric constraints. Solving the structures, via crystallography, with the thermodynamics data will provide supporting evidence to better explain changes to structure, stability, and function.

Initially, the R55 position was a primary interest because of the potential electrostatic interaction between R55 and D32. It was thought that the stabilizing effects coming from the interaction was significant for the stability and function of the protein. The results from the R55 library are still being collected, but the current results are listed in Appendix B (Kumar). The thermal and chemical melting temperatures are listed in Figure B1(a) and the *in vivo* activity is listed Figure B1(b). The two most stable mutants are R55L and R55I, with melting temperatures of 70.5^{o}C and 66.4 ^{o}C. Additionally, the activity of the Leucine and Isoleucine mutants is similar to wild-type Rop. Figure B2 shows the *in vivo* screen as developed by Magliery and Regan[17] and Figures B4 and B5 show the thermal and chemical melting values. Gibbs-Helmholtz analysis quantifies the conformational stability of a protein by measuring the change in Gibbs free energy from the folded state to the unfolded state. The ΔG of R55L was higher than R55, as shown in Figure B6, indicating that the conformational stability created by Leucine is more favorable than Arginine. The stability and activity of R55L and R55I directly contradicts the initially hypothesis for the R55 library that suggested the importance of the R55-D32 electrostatic interaction. However, previous studies have effectively repacked the core of Rop with Leucine and Isoleucine.[32,33,44] Both amino acids have favorable size constraints to pack within the limits of the 'd' position in the heptad packing pattern. The Ala_2Leu_2 variants all pack well and have thermal melting temperatures above or similar to wild-

type Rop. The Ala_2Leu_2-8 and Ala_2Leu_2-8-rev variants have the highest melting temperatures, but the activity of the variants were severely hindered. Likewise, the Leu_4 and Ala_2Ile_2-6 variants both had high melting temperatures, but they were in active according to the *in vivo* assay. The mutation of the Arginine out of the 55th position reduces favorable enthalpy contributions, but it is possible that the steric compatibility of Leucine and Isoleucine form favorable van der Waals interactions within the core and with Phe-56. The stability and activity of the other hydrophobic mutants (Met, Trp, Ala, Val) also suggest that the size compatibility of Leu and Ile optimize the van der Waals interactions and increase stability. Met has a similar structure to Leu, with the exception of the sulfur atom, but Trp is much larger than Arg and Val and Ala are much smaller. In addition, Lysine has a positive charge with a pKa near that of Arginine, but the slight size difference might be the cause for the lower melting temperature. Previously solved crystal structures show that the heptad packing pattern is broken at the 55th and 56th positions in that Phe-56 in the 'e' position packs in the core and Arg-55 in the 'd' position packs outside of the core. However, this hypothesis leaves a very interesting question, if Leucine or Isoleucine is mutated in the 55th position, does Phe-56 still take the place of the 'd' position in the core?

The secondary structure of the Rop variants was examined through circular dichroism. Circular dichroism is an effective method for detecting secondary structure and an α-helix creates a distinctive CD spectrum with negative peaks at 222 nm and 208 nm. The CD spectra of R55L and R55I were both similar to the wild-type spectrum, as seen in Figure B3, which indicates that both are helical structures. A ^{1}H-^{15}N HSQC was conducted to evaluate structural changes in relation to the structure of AV Rop. The results of the spectrum are shown in Figure B7. The R55L variant peaks are represented by the red dots and AV Rop peaks are in blue. The peak shifts caused by the Leucine in the 55th position are relatively minor and the spectrum closely resembles the spectrum of AV Rop. It appears that the hydrophobic Leucine in the 55th position does not significantly change the structure, function, or topology of the protein. The R55L variant has not been crystallized yet, but a solved crystal structure should provide more conclusive evidence about the packing effects Leucine.

The goal is to engineer variants with native-like stability and function so that terminal core packing impacts can be better understood. The focus of the study is to shed light on the functional roles of Arg-55 and Phe-56 and to find the governing interactions in those positions that dictate "native-like" stability and function. The results of the R55 library indicated that steric compliance and favorable van der Waals interactions could be achieved and overcome the necessity for an electrostatic interaction with D32. However, Phenylalanine is a large, aromatic and hydrophobic residue that might restrict the flexibility of the 55^{th} position. Previous studies of Rop have shown that unfavorably packing the core can detrimentally effect the stability and function of the protein. Creating a library in both the 55^{th} and 56^{th} positions might allow more flexibility in charge and packing compatibility between two amino acid pairs. It is thought that it might be possible to reverse the R55F56 sequence to F55R56 and correct the heptad repeat packing pattern by placing a hydrophobic amino acid back in the 'd' position. It is hypothesized that the electrostatic interaction is important to the stability of Rop and amino acid pairs that retain this interaction while also favorably filling the core space will have similar stability and function as AV Rop.

Chapter 2: Materials and Methods

2.1 Cloning

The library was constructed with a reengineered version of wild-type Rop, called AV Rop, that replaced C38 and C52 with Alanine and Valine.[48] The Rop gene was ordered in four fragments and reassembled through PCR techniques. The figure of the fragment is shown below in Figure 2.1. The R55F56 library was generated by mutating the codons at the 55th and 56th positions to NNK-NNK. The symbol 'N' indicates Adenosine, Guanine, Cytosine, or Thymine and the symbol 'K' encodes Guanine or Thymine. There were two reasons for mutating the third nucleotide of the codon sequence to K. First, it minimized the likelihood of a stop codon and second, it reduced the codon bias and thus more evenly distributed the statistic chance of any 1 of the 20 natural amino acids to be coded for. The library had the potential to produce 400 (20^2) possible variants, 20 possible amino acids at each position. The cloning procedure used Pfu DNA polymerase to reassemble the OP5 and OP3 fragments. For a standard 100 µL reaction, 1 µmol OP5 fragment, 1 µmol OP3 fragment, 1 µmol dNTPs, and 1 µmol Klenow Polymerase were mixed into 1x Neb2 buffer and the fragments were extended at 37 ^{0}C. The final product was mixed together with 1 µmol AP3, 1 µmol AP5, 1 µmol dNTP, and 1 µmol Pfu polymerase in a 25 µL reaction and PCR was carried out. The AP3, AP5, and product of the OP3/OP5 reassembly were mixed together in similar conditions with the addition of 0.5 µL of the OP5/OP3 product. The amplified Rop gene insert was digested with AflIII and BamHI and then cloned into the pACT7lacCm-AflIII vector.

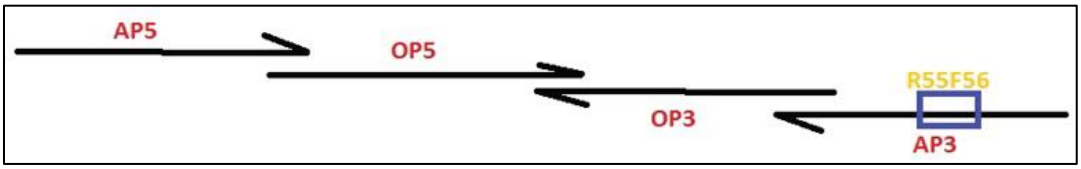


Figure 2.1: Illustration of the overlap and amplification primers that were ordered and then reassembled with PCR

2.2 Screening

The R55F56 library was transformed into DH10Bcells containing the pUCBADGFPuv screening plasmid, pUCBADGFPuv, developed by Magliery and Regan.[7] The screening plasmid enables active variants of Rop to be selected by linking ColE1 copy number to GFPuv expression.[7] GFPuv, which is a modified version of GFP, expressed via an arabinose promoter, with an 18 fold increase in fluorescence, and excitability by UV light at 395 nm.[7] After the R55F56 library variants were transformed they were recovered in 5 mL of 2YT at 37 ^{o}C and then plated on KAN+AMP+LB solid media plates that were infused with 0.0005% arabinose. The cells were incubated at 42 ^{o}C and grown for 16 – 18 hours. The cells were counted to ensure there was sufficient coverage of the possible 400 variants in the R55F56 library. The culture plates were then placed under UV-light and fluorescent colonies were picked from the selection plates.

2.3 Expression and Purification

Active variants were sent for colony sequencing to Genewiz. Selected variants were cloned into pMRH6TEV vector for expression using similar cloning scheme as above. The method for expression and purification of Rop was developed by Magliery in 2006. The variants were grown in 1L cultures of 2YT with KAN until an optical density of 0.8 at 600 nm was achieved. The cultures were then induced with 0.1 mM IPTG and grown for 16 – 18 hours at 30^{O}C. The cell cultures were centrifuged and the pellet was resuspended in Lysis Buffer (50 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). Detergents and enzymes were added to the solution to aid in cell lysis, the final solution concentrations are listed: 5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 25 μg DNase I, 150 μg RNase A, 0.1% Triton X-100. The cells are then lysed under very high pressure with an Emulsiflex. The Emulsiflexed solution was centrifuged, forming a pellet of cellular fragments and insoluble molecules at the bottom of the centrifuge tube. The soluble fraction is decanted off and saved for further purification. AV Rop is a soluble protein and all of the R55F56 library variants were primarily observed in the soluble fraction. The purification of Rop variants utilized six histidine residues to chelate nickel, in recyclable NiNTA agarose beads, which allowed impurities to be washed from the solution with Wash Buffer (50 mM phosphate, 300 mM NaCl, 20 mM imidazole, pH

7.4). The protein is then eluted from the column The NiNTA bound Rop protein is separated from the solution with a BioRad prefritted column using elution buffer (50 mMpPhosphate, 300 mM NaCl, 100 mM imidazole, pH 7.4. TEV protease was used to hydrolyze off the histidine residues at a TEV site that separated the binding tag and the Rop variant. The purity of the Rop variants was evaluated on an 18% SDS-PAGE gel (Polyacrylamide Gel Electrophoresis).

2.4 Thermal Denaturation

Spectra were obtained on a JASCO J-815 Circular Dichroism Spectrometer (The Ohio State University, Department of Chemistry). Rop variants were measured at 50 μM monomer in CD Buffer (50 mM sodium phosphate, pH 6.3, 300 mM NaCl). The concentration of Rop was calculated from UV absorbance at 280 nm and then quantified in an SDS gel against a standardized concentration of lysozyme. The samples were loaded into a 1 mm cuvette and the CD signal was measured at 25^{o}C from 260 nm to 190 nm. The scan was used to determine the overall helical structure of the variant. Thermal denaturation of the variants measured the mean residue ellipticity of the protein at 215 nm, 222 nm, and 600 nm wavelengths from 15 ^{o}C to 95 ^{o}C. The thermal melts were conducted at a rate of 1 ^{o}C/min with 1 ^{o}C step sizes and 6 seconds of equilibration at each temperature. The reverse melt, from 95 ^{o}C to 15 ^{o}C was collected as well and evaluated with the forward melt to ensure each variant was fully reversible. The melting temperature (T_m) was found by fitting the melting curve, in Microsoft Excel, to a Clark and Fersht equation used by Magliery Lab for previous experiments. $F = [(\alpha_F+\beta_F T) + (\alpha_U+\beta_U T) \exp((mT-T_m)/RT)] / [1 + \exp((mT - T_m)/RT)]$.[49] The α_F and β_F are the intercept and slope of the folded baseline. The α_U and β_U are the intercept and slope of the unfolded baseline. T is the temperature at that specific data point and T_m is the temperature at which 50% of the protein is unfolded. The T_m corresponds to 50% unfolded because the temperature denaturation of Rop is reversible. The reversibility of the protein samples were checked by looking for hysteresis between the forward and reverse thermal melt.

2.5 Chemical (Urea) Denaturation and Gibbs-Helmholtz Analysis

The spectra were obtained on the same equipment as the thermal denaturation experiments. Rop variants were measured at 50 μM monomer in CD Buffer (50 mM sodium phosphate, pH 6.3, 300 mM NaCl), and the concentration of Rop was calculated in the same manner as the thermal denaturation experiments. A stock of 10 M urea in CD buffer was made every time samples were prepared. A refractometer was used to determine the concentration of urea. Prepared samples were equilibrated for 48 hours at room temperature and then kept at 4 ^{o}C for 1 to 7 days before taking the spectra. The mean residue ellipticity of the 222 nm wavelength was measured at 25^{o}C in 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, or 5.5 M urea. The data was then fit to a Clark-Fersht equation, the same equation that is used to solve the melting temperature, except the temperature and melting temperature is replaced with the urea concentration and chemical melting concentration (M of urea). $F = [(\alpha_F+\beta_F[D]) + (\alpha_U+\beta_U[D]) \exp(m[D]-[D_{1/2}])/RT)] / [1 + \exp((m[D] - [D_{1/2}])/RT)]$. The data was fit with Microsoft excel with a least squares regression. The α_F, β_F, α_U, and β_U are explained above in the thermal denaturation equation. The D and $D_{1/2}$ is the concentration of urea and the concentration of urea at which 50% of the protein is unfolded. The chemical melting concentration is when the protein is 50% unfolded because the urea denaturation of Rop is reversible.

The purpose of the Gibbs-Helmholtz analysis was to measure ΔG, the change in free energy between the folded state and unfolded state, which provides information about the stability properties of the protein. A modified Gibbs-Helmholtz equation was derived in the introduction with the use of T_m as the reference temperature, however the study used T_g as the reference temperature parameter.[21] Becktel and Schellman were able to define three postulates in 1987 to derive a modified integrated form of the Gibbs-Helmholtz equation.[22] The denaturation reaction is defined as a two state process, $A_2 \leftrightarrow 2U$, the protein is stable at some temperature, and ΔC_P is equal to a constant greater than 0. Privalov and Khechinashvii showed that the third postulate was valid by showing that for a given protein, ΔC_P can be assumed constant within experimental error.[50] The equation Becktel and Schellman derived combined the following equations, $\Delta H_T = \Delta H_g + \Delta C_P(T - T_g)$ (1), $\Delta S_T = \Delta S_g + \Delta C_P \ln(T/T_g)$ (2), and $\Delta G = \Delta H - T\Delta S$ (3) to produce the Gibbs-Helmholtz equation $\Delta G_T = \Delta H_g (1 - T/T_g) - \Delta C_P(T_g - T) - \Delta C_P(T * \ln(T/T_g))$ (4),

ΔG_T is the change in Gibbs free energy as a function of temperature.[22,51] Gibbs free energy at the temperature of maximum stability, referred to as the conformational stability at T_s, is defined as $\Delta G_s = \Delta C_P (T_s - T_h)$.[22] T_s is the temperature of maximum stability and is defined by the equation $T_s = T_g / e^{[\Delta H/(\Delta C_P T_g)}$.[22,51] T_g is the temperature when $\Delta G = 0$ and from the standard Gibbs free energy equation, $\Delta G = \Delta H - T\Delta S$, the following relationship is formed: $\Delta S_g = \Delta H_g / T_g$.

The thermal denaturation curves were collected on the samples that were prepared and measured for the standard urea melts. The samples were heated from 5 ^oC to 85 ^oC and the CD signal was measured from 260 nm to 190 nm at each temperature. The temperature rate change was 5 ^oC per minute, scanning speed 100 nm/min, a bandwidth of 1 nm, and a 5 minute equilibration time preceded each measurement. The data was normalized for each temperature and the equilibrium constant for unfolding was calculated $K_d = (F_U^2/F_F)*2(50 \times 10^{-6} M)$. The fraction folded, F_F, and fraction unfolded, F_U, are related through the two state transition process and the dimerization of the protein is concentration dependent. An initial ΔG value was calculated from the K_d, by the equation $\Delta G = -RTlnK_d$. The ΔG was graphed and data points that were not in the linear transition range were deleted from the data set. The error associated with the ΔG calculation increases as the distance from the point of 50% fraction unfolded increases, so generally ΔG values are only calculated from data points associated with the unfolding transition of the protein. The ΔG_{H2O} of each temperature was calculated by the equation $\Delta G_u = \Delta G_{H2O} + m[\text{urea}]$, the change in ΔG is linearly dependent upon the concentration of urea.[52] The temperatures found in the unfolding transition of the protein at 0 M urea were not used to calculate ΔG_{H2O}. The temperatures found in the unfolding transition of the protein at 0 M urea were used to calculate ΔG through the equation, $\Delta G = RT*\ln(2(50/1000000)*(F_U^2/F_F)/1000$. The fraction folded and fraction unfolded were corrected for baseline shifts. A sum of squares was then used to minimize the error and create a best line of fit for ΔG, with the parameters ΔH, ΔC_P, T_g, ΔS, T_s, and ΔG_s.

2.6 ^1H-^{15}N HSQC NMR

The HSQC spectra were collected on a Bruker DMX-600 MHz spectrometer. The expression and purification procedures for the ^1H-^{15}N HSQC were similar to that of the

standard expression and purification protocols, with only a few minor changes. The variants were expressed in 1 L of minimal media (100 mM Na-K (1:2) phosphate, 5 mM sodium citrate, 0.5 mM $MgSO_4$, 0.6% glucose, 10 mgL^{-1} thiamine, 0.3 % trace metals (per liter dissolved with 10% HCl, 10 g $FeCl_3$-$6H_2O$, 2 g $ZnCl_2$-$4H_2O$), 2 g $CoCl_2$-$6H_2O$, 2 g Na_2MoO_4-$2H_2O$, 1 g $CaCl_2$-$2H_2O$, 1 g $CuCl_2$, 1 g $MnCl_2$, 0.5 g H_3BO_3, pH 7.2) and addition of 1 g of [^{15}N]-NH_4Cl as the nitrogen source. The protein was in CD buffer (50 mM sodium phosphate, 300 mM NaCl, pH 6.3) with 10% D_2O and concentrated to 0.5 – 2 mM. The experiment was conducted at 25^{o}C.

Chapter 3: Results and Discussion

3.0 Contributions

The majority of the work presented for the R55F56 library was conducted and analyzed by the primary author. All of the work presented for the R55 library was conducted and analyzed by Anusha Kumar. All of the data for the R55 library was kept separate in Appendix B and only referenced in this senior thesis because it helped explain the molecular interactions occurring at the 55th and 56th positions.

3.1 Screening R55F56 Libraries

The R55F56 library randomized codon sequence at the 55th and 56th positions to NNK-NNK. The N indicated any four of the nucleotides (Thymine, Adenine, Cytosine, Guanine) and the K indicated either Guanine or Thymine. The purpose of randomizing the codon sequence with the K was to minimize the likelihood of a stop codon and more evenly distribute the statistic chance of any 1 of the 20 natural amino acids to be coded for. The library was able to produce 400 (20^2) possible variants, so to ensure enough variants were screened, variants were assayed by the *in vivo* screen developed by Magliery and Regan and then sequenced until repeats were prevalent. The frequency of amino acids at the 55th and 56th positions were listed below in Figure 3.1.

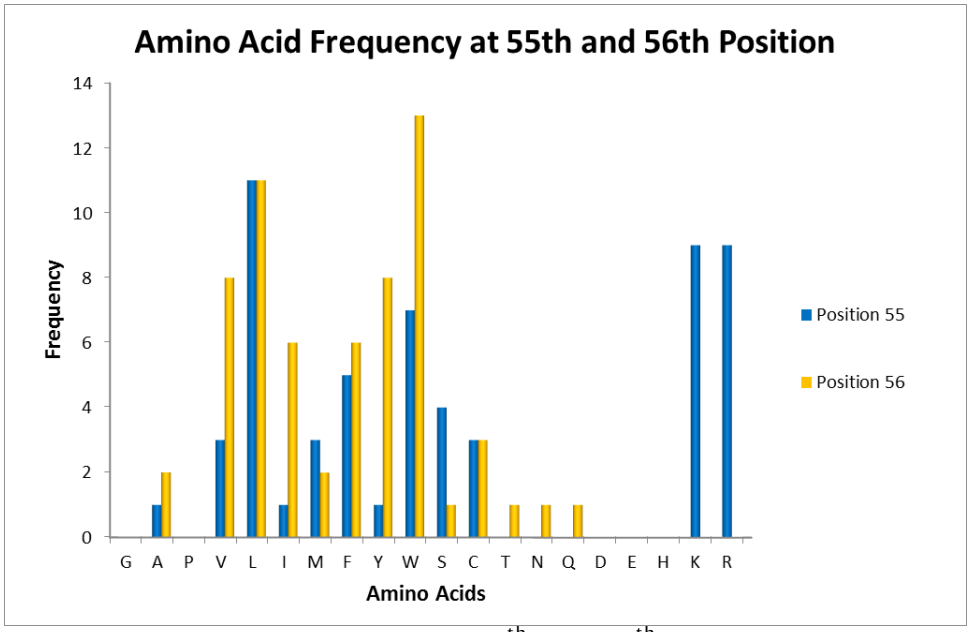


Figure 3.1: Amino Acid Frequency at 55th and 56th Positions

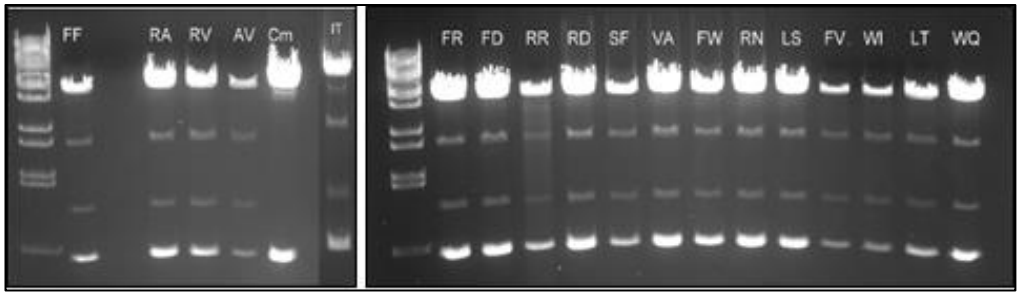
It was intuitively expected that the number of positively charged amino acids at the 55th position would have a rather high frequency. The current hypothesis suggests that the positive charge of R55 interacts with the negative charge found on D32, which is in the loop region of the opposite monomer. Lysine, K, is slightly smaller than the Arginine, which is found in native Rop, but could still potentially fill the role of the charge interaction with D32. However, hydrophobic amino acids were also very common in both the 55th and 56th positions, which contradicts the hypothesis that the R55-D32 interaction is an essential stabilizing electrostatic interaction. Some structural papers have suggested that the Phe56 plays a role of a hydrophobic plug at the terminal end of the core.[28,33] Packing a hydrophobic residue into the core could form favorable van der Waals interactions and limit the entropic destabilization that would occur if more of the hydrophobic core was exposed to the aqueous solution.

The *in vivo* functional assay, developed by Magliery and Regan[17] links the functional abilities of the Rop mutants to the fluorescence of GFP. Rop naturally regulates the plasmid copy numbers of plasmids with a ColE1 origin within *E.coli*. Rop mutants that function similar to that of wild-type Rop limit the plasmid copy number of the plasmid that contains the GFP gene. It could be assumed that the level of GFP expression, and thus fluorescence, was directly proportional to the amount of plasmid that contained the GFP gene. Therefore, the mutants that produced Rop variants with native-like function would be dim and inactive variants would express high levels of GFP because the copy number of the plasmid would not be regulated. The problem with a negative screen is that mutations within the GFP gene or other mutations could have downstream effects that limited the fluorescence of the GFP and potentially creating false positives. The positive screen created by Magliery and Regan utilizes the AraC gene with an arabinose promoter. Growing the cells on an LB medium with 0.0005% arabinose is enough to turn the AraC gene into a promoter for cells that contain active Rop variants. The cells with inactive Rop variants have a much higher copy number of the pUCBADGFPuv plasmid and the low concentration of arabinose is not enough to overcome the inhibition effects of the AraC gene. Therefore, the positive screen generated fluorescence as a result of the activity of the Rop variants.

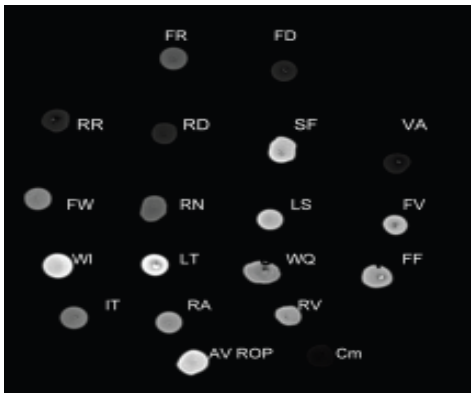
The results of the positive screen for the R55F56 library are shown below in Figure 3.2. Figure 3.2(a) shows the quantification of plasmid DNA from each of the Rop variants. The variants were grown under identical conditions at 42^{o}C and the optical density of the cultures were standardized so that each sample contained approximately the same number of cells. The cells were lysed with an alkaline lysis and the plasmid DNA was collected through the use of a Quiagen mini-prep kit. The plasmids were cleaved with AflIII and Bam HI, the top and bottom bands in the gel are fragments of the pUCBADGFPuv plasmid and the middle two bands are from the pAC plasmid carrying the Rop gene. Rop variants with native-like function regulate the plasmid copy number of the pUCBADGFPuv plasmid since it has a ColE1 origin. The positive band is labeled (AV), which is a cysteine free variant of wild-type Rop that has very similar functional capabilities. The negative band is labeled Cm, chloramphenicol, is a linker sequence that does not contain the Rop gene and therefore the plasmid copy number of the pUCBADGFPuv is not regulated. The results of the plasmid copy number test combined with the results of the *in vivo* assay, shown in Figure 3.2(b), were used to determine the activity of the Rop variant. The activity of the variants and is listed from most active to least active in Figure 3.2(c). The results of the screen indicate that mutants with hydrophobic amino acids are much better tolerated than variants with polar and charged amino acids.

The screen indicated that the FV and WI variants had native-like Rop activity. The removal of the positively charged Arginine eliminates the electrostatic interaction between R55 and D32. The variants with intermediate activities contain combinations of polar or charged amino acids and hydrophobic amino acids. The FR mutant conserves the heptad repeat pattern by packing the hydrophobic Phe in the ‘d’ position of the core and the positively charged Arg in the ‘e’ position. However, it appears that the reversal of the wild type packing pattern severely diminishes the functional capabilities of Rop. The inactive variants primarily contained amino acids with charge, with the exception of the Val-Ala variant. The VA mutant is supposedly very underpacked, compared to the volume of Phe + Arg in wild type Rop. The RD and FD variants were inactive, which was expected because of the unfavorable like charge repulsion that most likely occurred between D56 and D32.

(a)



(b)



(c)

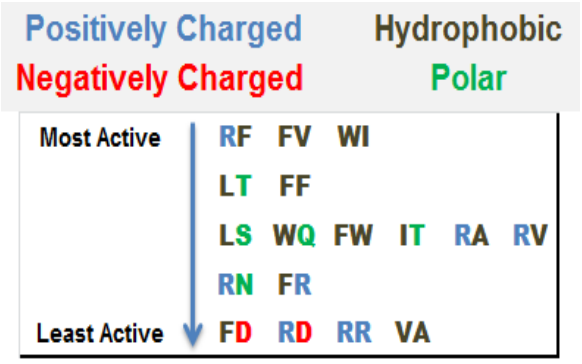


Figure 3.2: Screening results from the R55F56 Library. Figure (a), shows the an analytical digest of the pUCBADGFPuv and pAClacRop plasmids, the bottom and top bands are the pUCBADGFPuv plasmids with a ColE1 origin and the copy number is regulated by active Rop variants. Figure (b), shows the fluorescence of the GFP as a direct indictor of Rop function. (c) The color code is as follows: hydrophobic (brown), polar (green), positively charged (blue), negatively charged (red), special case (black).

Circular dichroism spectroscopy was used to analyze the secondary structure of the variants. Wild-type Rop is completely α-helical with the exception of the loop and last couple amino acids at the C-terminus. The CD signal of AV Rop at 50 μM had a recorded mean residue ellipticity of 20,000 deg cm^2 $dmol^{-1}$. Since the native structure of Rop is almost completely α-helical, no variant will produce a higher CD signal, within the experimental error of the concentration estimation. Since Rop is a homodimer, the formation of the homodimer is concentration dependent. A higher concentration of Rop will cause more dimer structures to form, which can artificially influence the CD signal. The CD spectrum was recorded from 260 nm to 190 nm, which included the 222 nm and 208 nm peaks that are commonly associated with α-helical secondary structure, and the CD spectrum can be seen below in Figure 3.3. The RR variant and the VA variant were the only two Rop mutants within the R55F56 library that were significantly less helical than wild-type Rop. The CD spectroscopy of the RN variant indicated higher α-helical

content than wild-type Rop, however as was discussed above, wild-type Rop is almost completely helical and the spectrum could have been influenced by concentration.

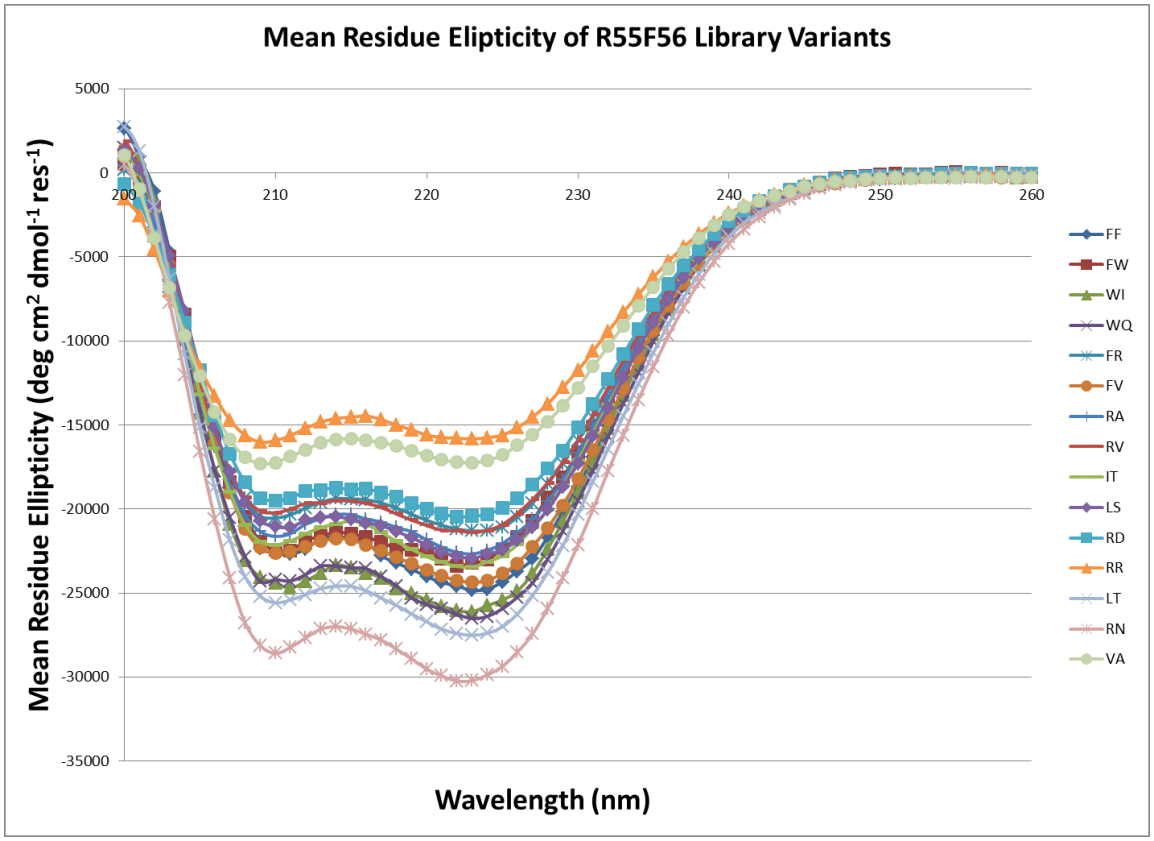


Figure 3.3: R55F56 Library Variants CD Scan of Mean Residue Ellipticity (MRE)

Thermal and chemical denaturation was monitored by CD, which detected the denaturation of the protein by monitoring the breakdown of the secondary structure. The results of the thermal denaturation experiments are listed below in Figure 3.4. AV Rop with Arg-55 and Phe-56 was the most stable variant found in the R55F56 library, and has a T_m of 66.81. The variants that contained large, aromatic hydrophobic amino acids in both the 55^{th} and 56^{th} positions are represented by the green data point in the figure. The two large aromatic amino acids remove any potential electrostatic interaction and pack a larger surface area and volume than AV Rop. The variants represented by blue data points are similar to the variants in green, with the exception that one of the amino acids is not aromatic. The required volume for WI and FV is slightly smaller than that required for FF and FW. The variants represented by the purple data points contain one

hydrophobic amino acid and one polar amino acid. The VA variant is represented by red data points and is separate from the other hydrophobic pairs because the packing volume is much smaller than the packing volume of AV Rop. The variants represented by the green and orange data points contain one or two amino acids with charge. The variants with higher melting temperatures contained primarily hydrophobic amino acids. The 55th and 56th positions appear to yield similar packing results as previous repacking studies in that overpacked variants are well accommodated for and underpacked variants are much more perturbed.[32,33] Another interesting result from the thermal denaturation data was the significantly lower melting temperatures of variants RV, RA, and FR. The RV and RA variants retain the electrostatic interaction with D32, but have a much smaller hydrophobic residue to pack the 'd' position of the core. Reversing the packing of the heptad repeat with the FR variant significantly lowered the melting temperature. The difference between the RF and FR thermal denaturation curves is shown in Figure 3.5. The results of the chemical denaturation experiments are similar to that of the thermal results, in respect to the stability of the protein. The urea melts are shown in Figure 3.6 and the thermal and chemical melting temperatures are listed in Table 3.1.

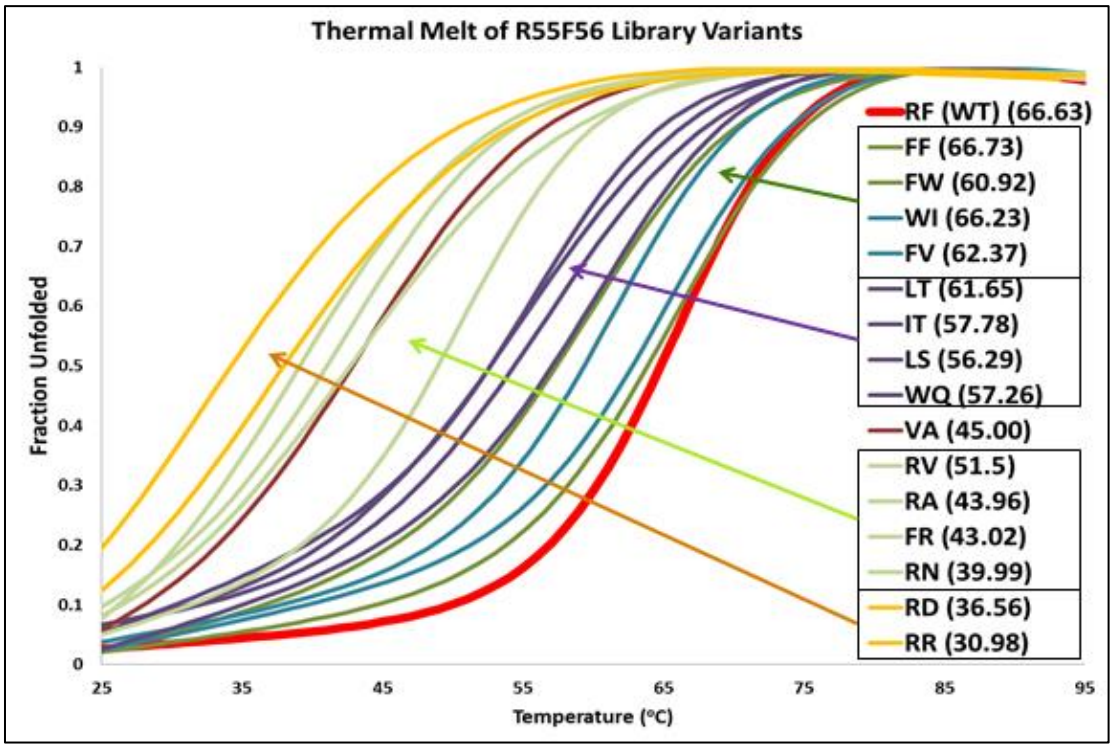


Figure 3.4: Thermal denaturation of the R55F56 Library Variants.

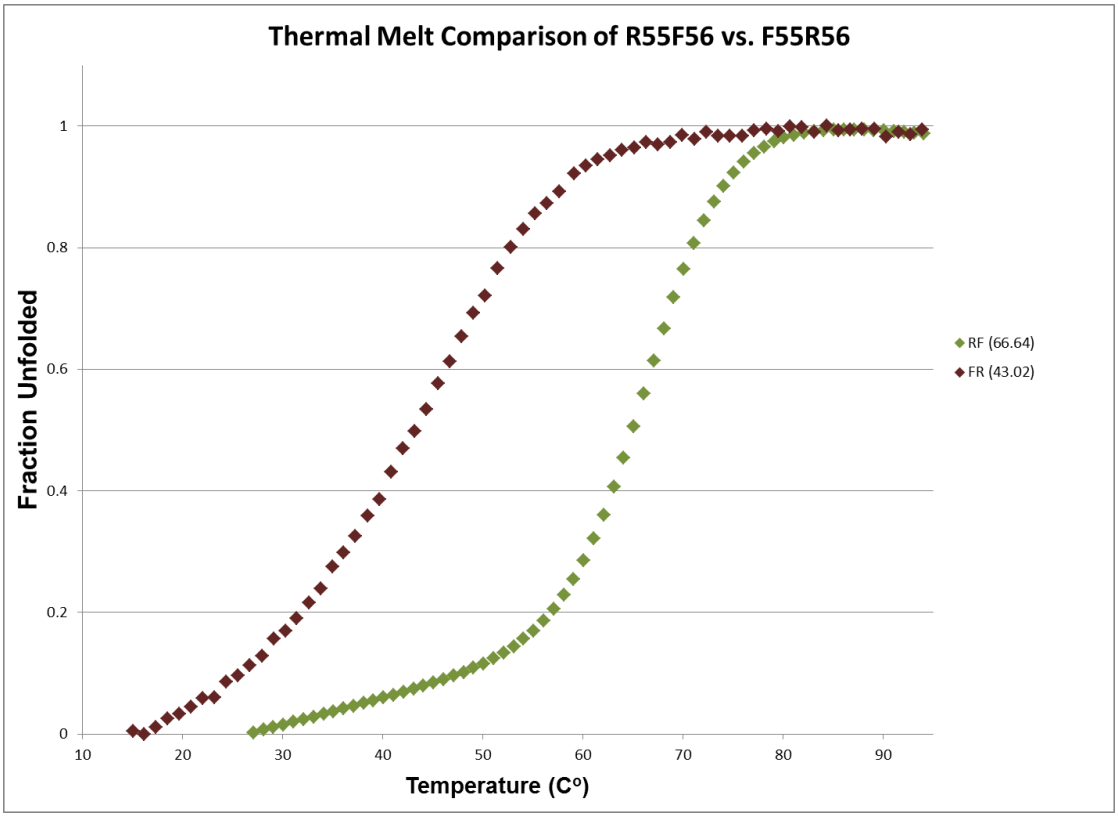


Figure 3.5: Thermal Denaturation of AV Rop (R55F56) versus Reversed Mutant (F55R56). Attempts to realign the heptad repeat packing pattern and restore the "d" position in the core.

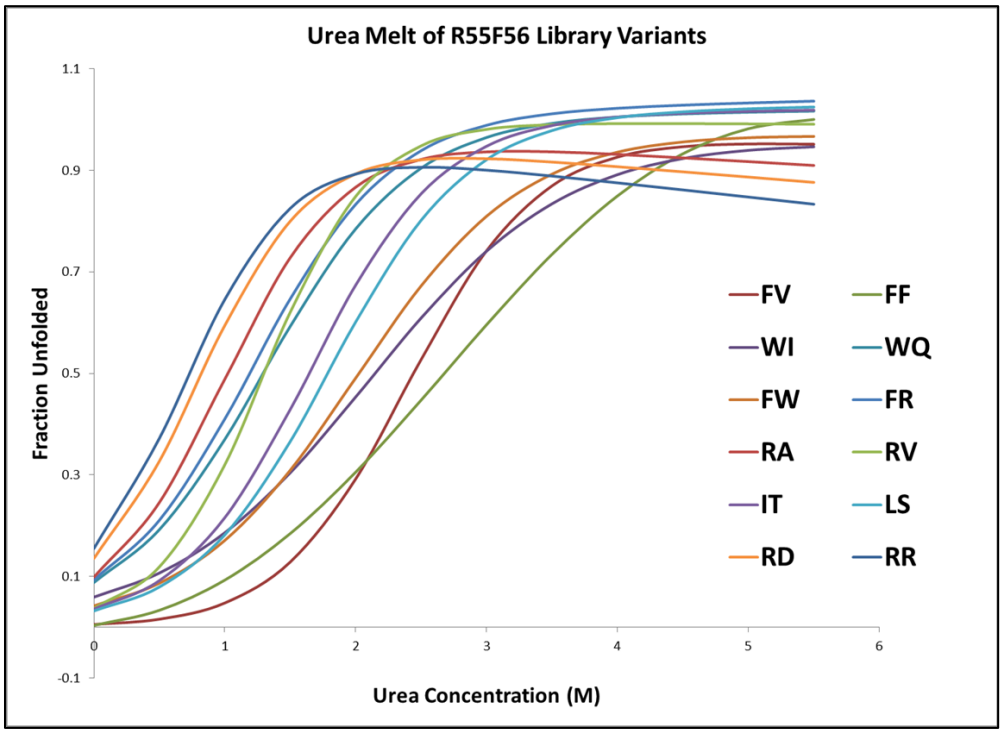


Figure 3.6: Chemical Denaturation with Urea of R55F56 Library Variants

R55F56 Variants	T_m °C	$D^{1/2}$ (M)
RF (Wild Type)	66.81	4.3
FF	66.73	2.77
FW	60.92	2
WI	66.23	2.1
FV	62.37	2.4
LT	61.65	-
IT	57.78	1.65
LS	56.29	1.8
WQ	57.78	1.3
VA	45	-
RV	51.5	1.3
RA	43.96	1
FR	43.02	1.2
RN	39.99	-
RD	36.56	0.8
RR	30.98	0.7

Table 3.1: R55F56 Library Variants Activity, Tm, and D1/2. T_m, melting temperature, $D^{1/2}$, is the chemical melting temperature with Urea.

Gibbs-Helmholtz analyses were conducted on interesting variants that helped explain the stabilizing effects of hydrophobic packing and the destabilizing effect of charged and polar residues. Figure 3.7 compares the conformational stability of AV Rop, LS, WQ, and RR. The RR variant is drastically less stable than AV Rop, Table A1 and A3 (Appendix A) show that the RR variants is less stabilized enthalpy and entropy. The close proximity of the two positive charges most likely contributes to the instability of RR. Additionally, the results of the R55F56 library show that van der Waals interactions and other favorable hydrophobic interactions play a large role in the stability of the 55 and 56 positions. Figure 3.8 shows the conformational stability of variants with pairs of hydrophobic amino acids. The FV variant has a slightly higher conformational stability than the AV Rop, despite the loss of positive charge. In addition, larger hydrophobic pairs that overpack the native volume, variant FF, are only slightly less stable than AV Rop. Further support of the importance of hydrophobic packing in the 55^{th} and 56^{th} positions can be seen by evaluating the conformational stability of the RA variant. The RA variant only changes the size of the hydrophobic residue from Phe to Ala, but the effects of the smaller Alanine are significant.

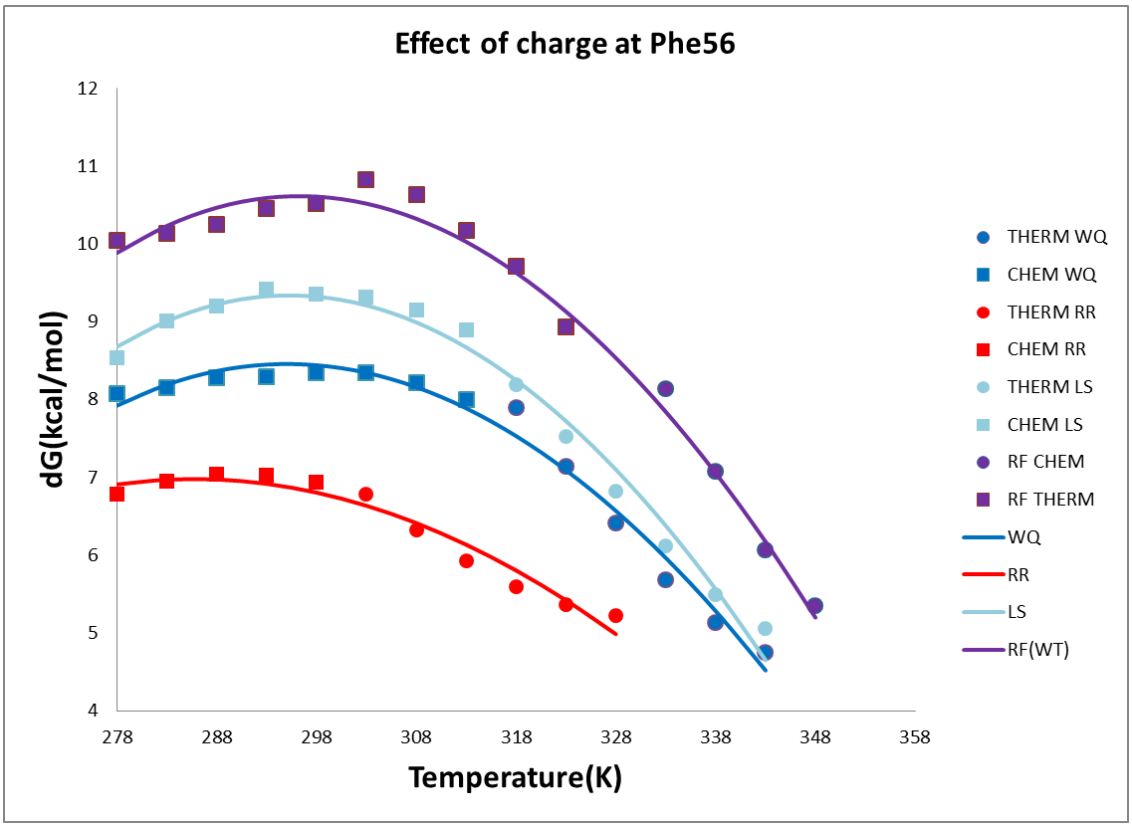


Figure 3.7: Gibbs-Helmholtz Effect of Charge at Phe56

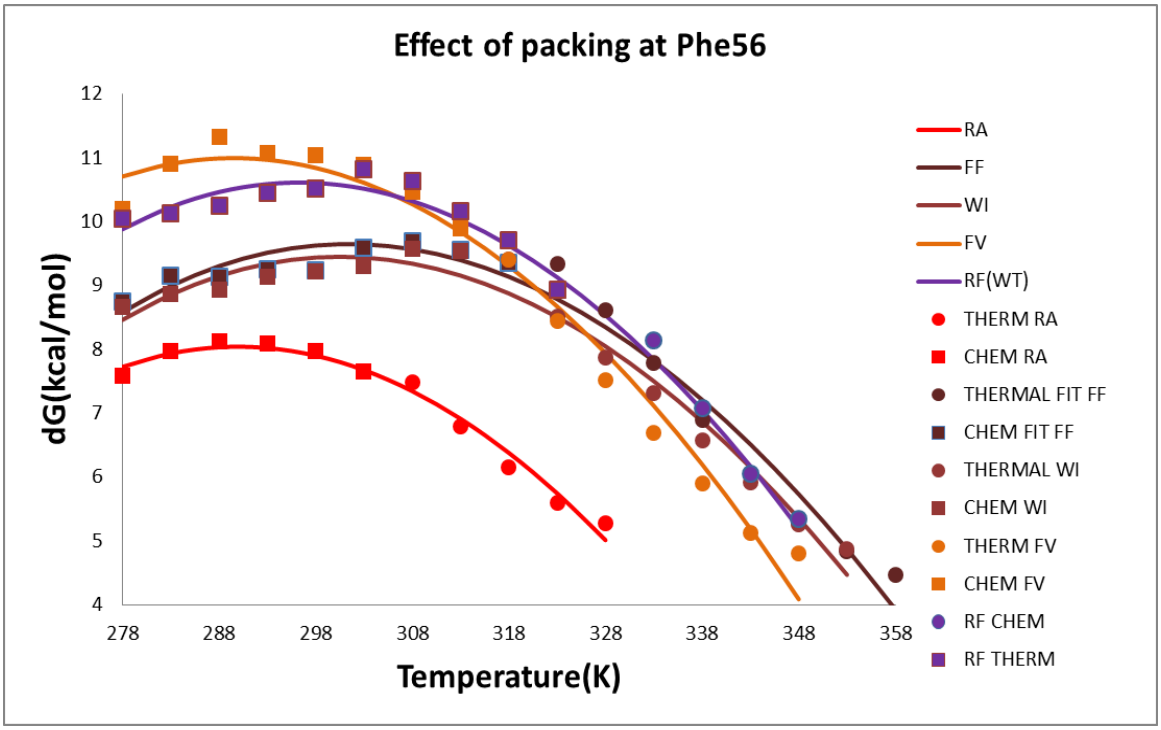


Figure 3.8: Gibbs-Helmholtz Effects of Packing at Phe56

The structural evaluations of the variants FF, FR, FV, LT, RN, and WQ are shown below in Figures 3.9 – 3.14, which contain ^{1}H-^{15}N HSQC peak shift analyses. All of the variants measured contained significant peak shits, which indicated that all of the structures had been slightly altered. The results were slightly unexpected because the variants only contained four amino acid mutations, two per monomer, and retained at least 97% of the wild-type sequence.

We were intrigued by the amount of peak shifts between variants such as FF and FR, shown in Figures 3.9 and 3.10. It was hypothesized that the FF variant would retain a native-like structure and that the drastic differences in the packing pattern/volume, stability, and *in vivo* function between the FF and FR variants would yield different peak shift patterns. However, the magnitude and variability of peak shifts between the two variants was indistinguishable. The FR variant repacks the 'd' and 'e' positions so that the hydrophobic and polar trends of the heptad repeat packing pattern were restored. In addition, the variant retains the same packing volume and surface area as the native RF packing structure. However, the FR variant is significantly destabilized and has lost the majority of its *in vivo* activity. The FF variant contains two large hydrophobic Phenylalanine amino acids, which produces a larger packing volume and surface area than the native RF packing structure. The two hydrophobic amino acids are not able to form an electrostatic interaction with D32 and therefore the variant must be compensating for the enthalpic penalty with favorable van der Waals interactions. Even though the loss of the electrostatic interaction did not cause a decrease in stability, it might have lessened the variants specificity to one particular conformation. A change in the structural conformation would change the binding affinity to RNA, which would result in the loss of *in vivo* activity. The FF variant produced *in vivo* activity levels very similar to that of wild-type Rop. Therefore, we think that the loss of the electrostatic interaction contributes to a higher conformational flexibility, which caused the observed peak shifts. Once the RNA loop comes into close proximity to the binding site the non-specific conformation, quickly flipping between atomic level conformational changes, becomes specific and binds RNA with similar affinity to that of wild-type Rop. From previous studies we know that the conformation of the variant must be similar to the wild-type variant and the perturbations we are suggesting are on a small atomic scale.[17]

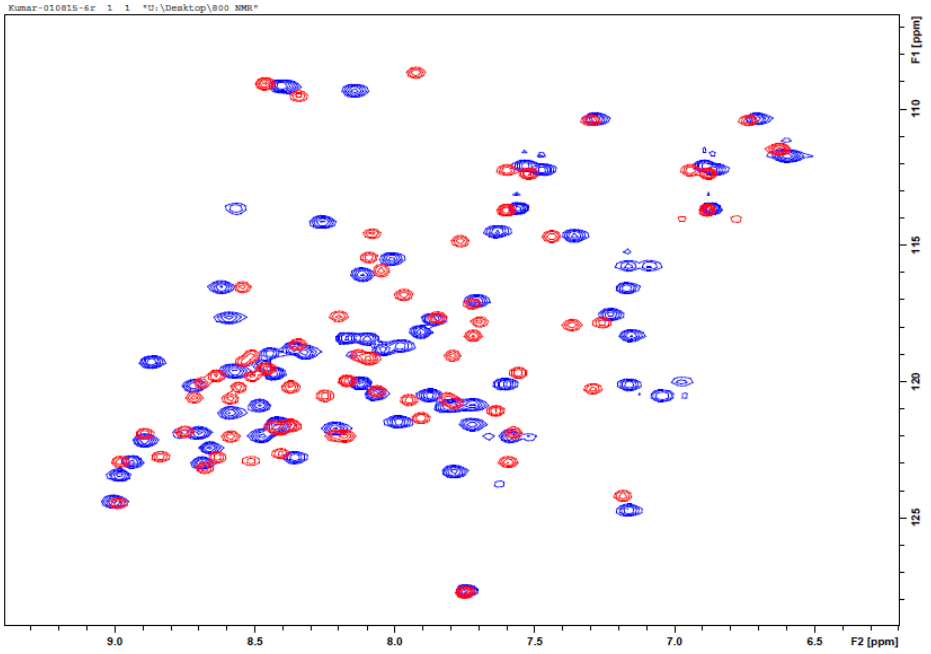


Figure 3.9: ^{1}H-^{15}N HSQC-NMR R55F56 Variant FF. Packing Phe at both the 55th and 56th positions creates a larger hydrophobic surface area and volume than native Rop.

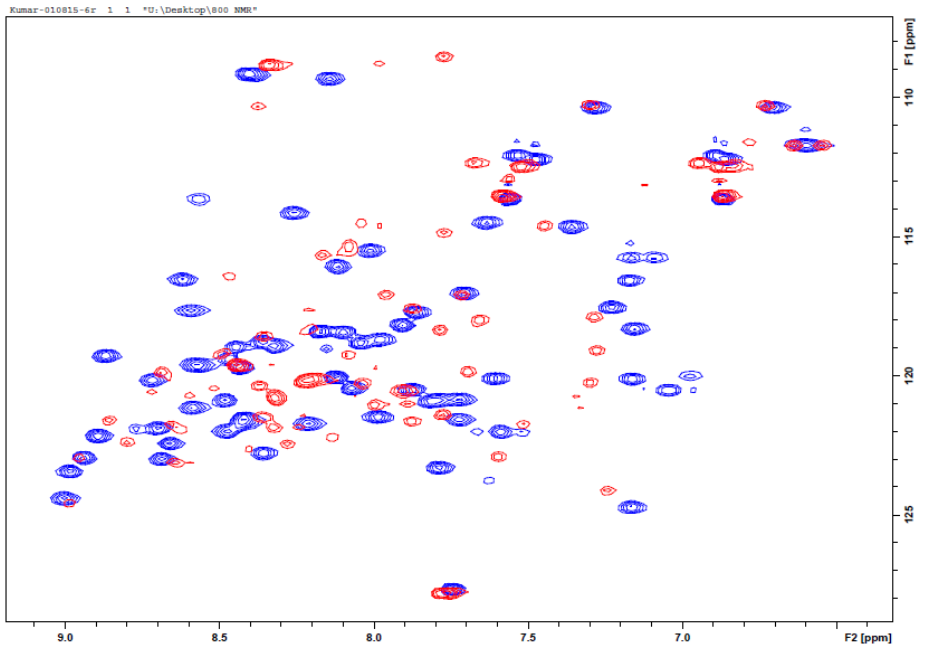


Figure 3.10: ^{1}H-^{15}N HSQC-NMR R55F56 Variant FR. Packing Phe at 55 and Arg at 56 reverses the native packing of Rop and restores a hydrophobic core packing residue to the "d" position of the heptad repeat pattern.

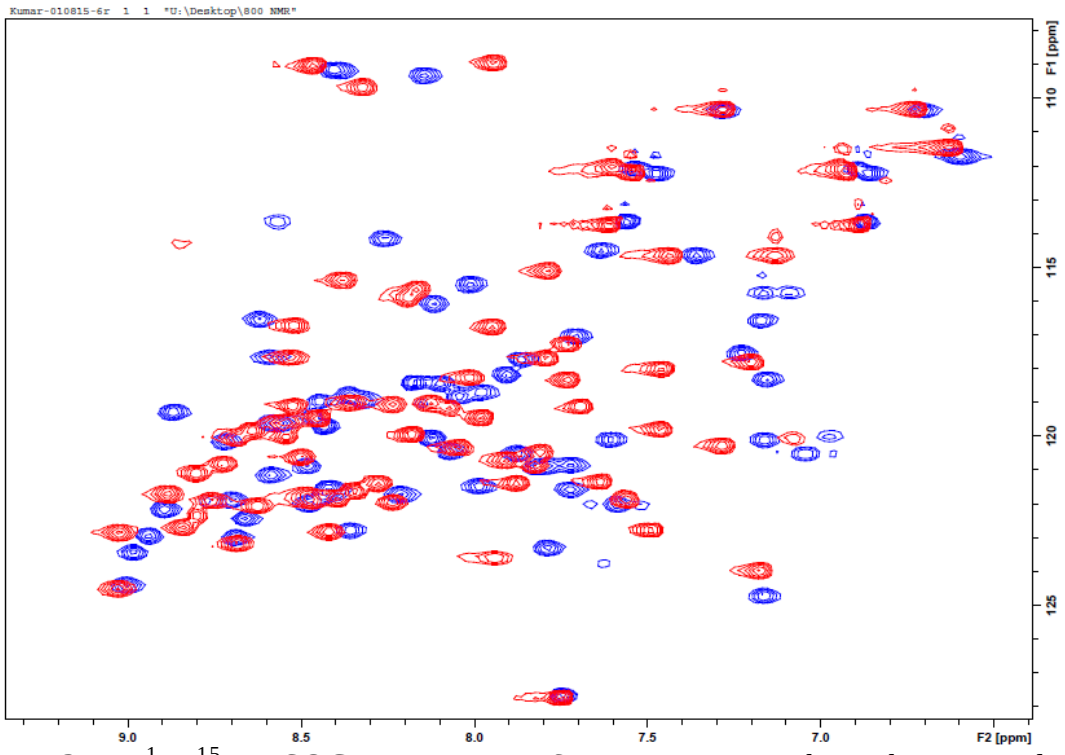


Figure 3.11: ^{1}H-^{15}N HSQC-NMR R55F56 Variant FV. Packing Phe at 55 and Val at 56 removes the possible electrostatic between R55-D32 in native Rop and inserts a larger, Phe55, at the "d" position, but the overall volume of F55V56 is slightly smaller than native Rop.

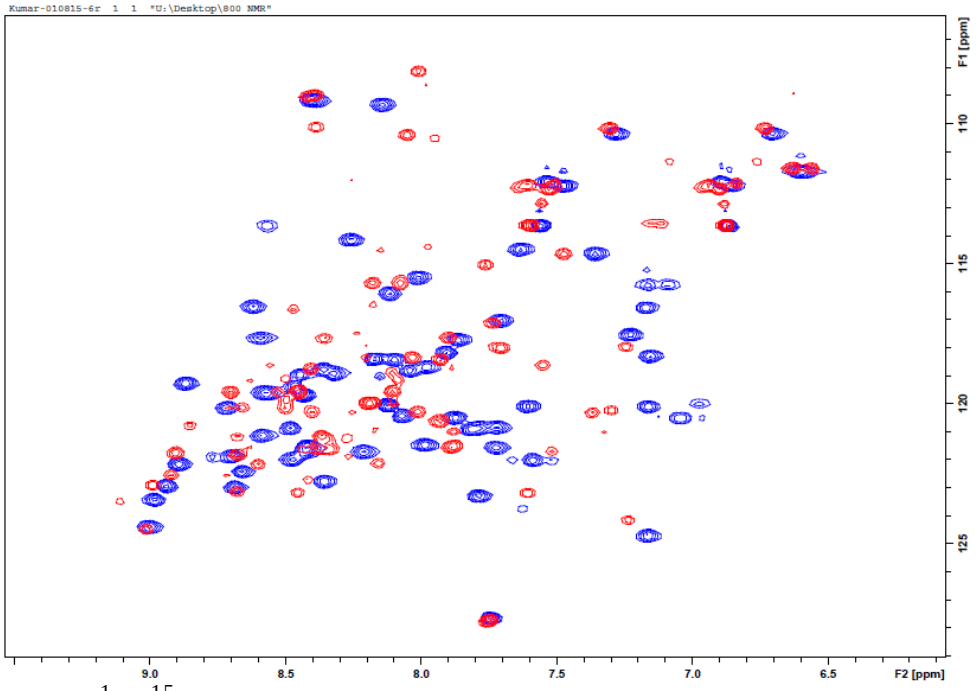


Figure 3.12: ^{1}H-^{15}N HSQC-NMR R55F56 Variant LT. Leucine has been shown to pack well at the "d" position in the Rop core and Thr at 56 introduces polar properties.

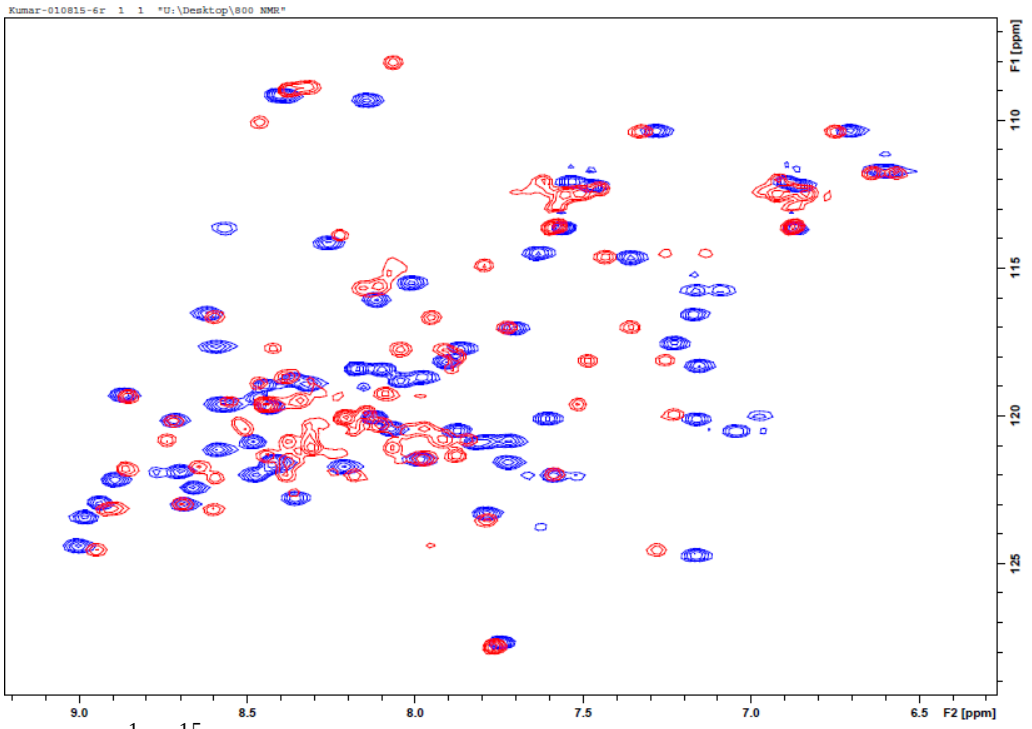


Figure 3.13: ^{1}H-^{15}N HSQC-NMR R55F56 Variant RN. Packing Asn at 56 severely limits the hydrophobic interactions that Phe56 fulfills in native Rop.

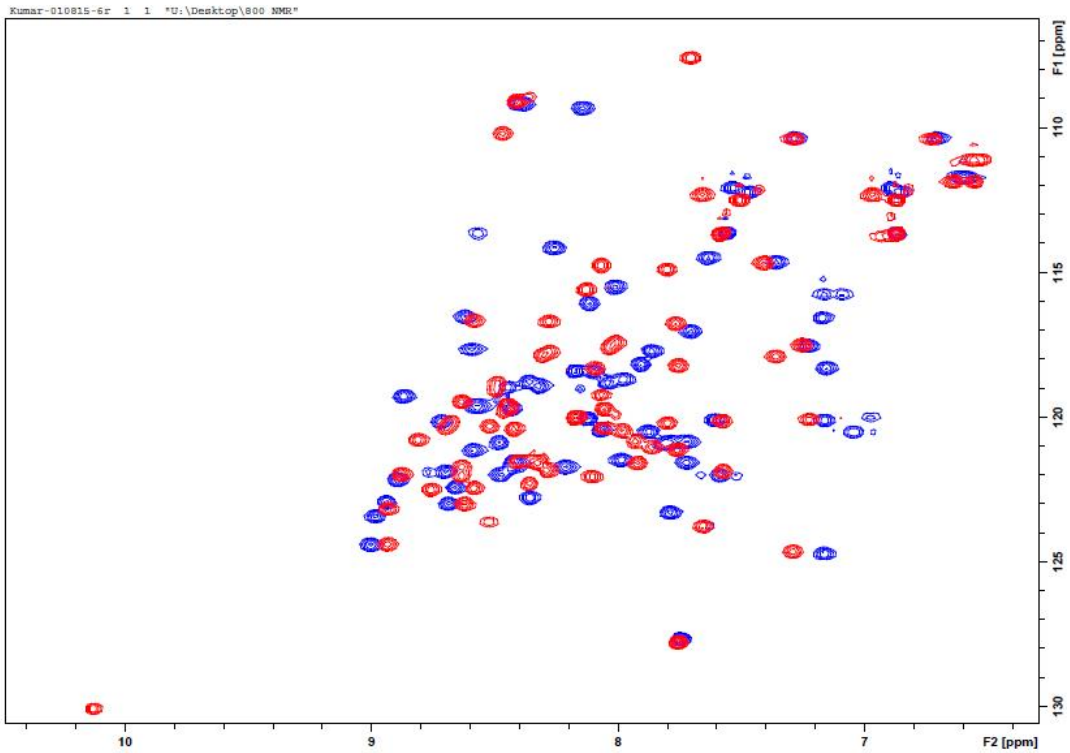


Figure 3.14: ^{1}H-^{15}N HSQC-NMR R55F56 Variant WQ. **Figure:** WQ Packing Trp at 55 introduces a large aromatic hydrophobic residue in a “d” position and Gln at 56 is a polar residue.

Chapter 4: Summary

The importance of the hydrophobic interactions and size packing constraints are evident in the results of the characterization of the R55F56 library. Variants that consisted of hydrophobic pairs and eliminated the electrostatic interaction with D32, were more stable than variants with charged amino acids. Hydrophobic pairs that overpacked the native volume were slightly destabilized and the *in vivo* function was hindered marginally. The hydrophobic pair that underpacked the native volume, mutant VA, reduced the stability of the protein more significantly and little to no *in vivo* activity was retained. Even when Arginine was conserved in the 55^{th} position, replacing Phenylalanine in the 56^{th} position with a smaller hydrophobic amino acid, such as Alanine or Valine, significantly destabilized the conformational stability of the protein. Some *in vivo* activity was still retained even though the 56^{th} position was underpacked from the native volume. The 56^{th} position was slightly destabilized by polar amino acids, but it was not able to support charged amino acids. Variants FF and FV, removed the electrostatic interaction with D32 and the stability was derived from favorable van der Waals packing. The variants LS, LT, IT, and WQ were destabilized relative to AV Rop and the *in vivo* function was impaired, but still maintained an intermediate activity. The variant FR, was an attempt to correct the heptad packing pattern by inserting a hydrophobic amino acid in the core 'd' position and moving the charged amino acid into the 'e' position. Reversing the native packing pattern in the 55^{th} and 56^{th} positions severely destabilized the protein and significantly hindered *in vivo* function. The protein appeared to still be folded, the CD spectrum still indicated an α-helical structure and the ^{1}H-^{15}N HSQC indicated the peak shift was still well dispersed. The destabilization of the FR variant suggested that the hypothesized importance of the electrostatic interaction with D32 was not an essential source of stabilization and only adds to the stability and function of the wild type variant because the charge can be tolerated in the 55^{th} position. Additionally, the instability and functional limitations that result from a polar, and especially charged residues, in the 56^{th} position indicates that the 56^{th} position serves as a hydrophobic plug to the terminal end of the core. Further structural analysis with crystallography should help clarify the structural and hydrophobic constraints of the 55^{th} and 56^{th} positions.

Bibliography

1 Anfinsen, C. B. The formation and stabilization of protein structure. *The Biochemical journal* **128**, 737-749 (1972).

2 Khoury, G. A. *et al.* WeFold: a coopetition for protein structure prediction. *Proteins* **82**, 1850-1868, doi:10.1002/prot.24538 (2014).

3 Baker, D. Centenary Award and Sir Frederick Gowland Hopkins Memorial Lecture. Protein folding, structure prediction and design. *Biochemical Society transactions* **42**, 225-229, doi:10.1042/BST20130055 (2014).

4 Boucher, R. C. An overview of the pathogenesis of cystic fibrosis lung disease. *Advanced drug delivery reviews* **54**, 1359-1371 (2002).

5 Kim, P. S. & Baldwin, R. L. Intermediates in the Folding Reactions of Small Proteins. *Annual review of biochemistry* **59**, 631-660, doi:DOI 10.1146/annurev.biochem.59.1.631 (1990).

6 Matthews, C. R. Pathways of protein folding. *Annual review of biochemistry* **62**, 653-683, doi:10.1146/annurev.bi.62.070193.003253 (1993).

7 Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annual review of biophysics* **37**, 289-316, doi:10.1146/annurev.biophys.37.092707.153558 (2008).

8 Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042-1046, doi:10.1126/science.1219021 (2012).

9 Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M. & Hecht, M. H. Protein Design by Binary Patterning of Polar and Nonpolar Amino-Acids. *Science* **262**, 1680-1685, doi:DOI 10.1126/science.8259512 (1993).

10 Ponder, J. W. & Richards, F. M. Tertiary Templates for Proteins - Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes. *Journal of molecular biology* **193**, 775-791, doi:Doi 10.1016/0022-2836(87)90358-5 (1987).

11 Hutchison, C. A. *et al.* Mutagenesis at a Specific Position in a DNA-Sequence. *J Biol Chem* **253**, 6551-6560 (1978).

12 Sigal, I. S., Harwood, B. G. & Arentzen, R. Thiol-Beta-Lactamase - Replacement of the Active-Site Serine of Rtem Beta-Lactamase by a Cysteine Residue. *P Natl Acad Sci-Biol* **79**, 7157-7160, doi:DOI 10.1073/pnas.79.23.7157 (1982).

13 Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. & Smith, M. Redesigning Enzyme Structure by Site-Directed Mutagenesis - Tyrosyl Transfer-Rna Synthetase and Atp Binding. *Nature* **299**, 756-758, doi:Doi 10.1038/299756a0 (1982).

14 Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-126, doi:10.1038/340122a0 (1989).

15 Matthews, B. W. Genetic and Structural-Analysis of the Protein Stability Problem. *Biochemistry* **26**, 6885-6888, doi:Doi 10.1021/Bi00396a001 (1987).

16 Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nature reviews. Molecular cell biology* **3**, 964-970, doi:10.1038/nrm975 (2002).

17 Magliery, T. J. & Regan, L. A cell-based screen for function of the four-helix bundle protein Rop: a new tool for combinatorial experiments in biophysics. *Protein engineering, design & selection : PEDS* **17**, 77-83, doi:10.1093/protein/gzh010 (2004).
18 Magliery, T. J. & Regan, L. Combinatorial approaches to protein stability and structure. *European journal of biochemistry / FEBS* **271**, 1595-1608, doi:10.1111/j.1432-1033.2004.04075.x (2004).
19 Lin, H. & Cornish, V. W. Screening and selection methods for large-scale analysis of protein function. *Angew Chem Int Ed Engl* **41**, 4402-4425, doi:10.1002/1521-3773(20021202)41:23<4402::AID-ANIE4402>3.0.CO;2-H (2002).
20 Lavinder, J. J., Hari, S. B., Sullivan, B. J. & Magliery, T. J. High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J Am Chem Soc* **131**, 3794-3795, doi:10.1021/ja8049063 (2009).
21 Pace, C. N. & Laurents, D. V. A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-2525 (1989).
22 Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26**, 1859-1877, doi:10.1002/bip.360261104 (1987).
23 Itoh, T. & Tomizawa, J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc Natl Acad Sci U S A* **77**, 2450-2454 (1980).
24 Tomizawa, J. Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript. *Cell* **38**, 861-870 (1984).
25 Cesareni, G., Muesing, M. A. & Polisky, B. Control of ColE1 DNA replication: the rop gene product negatively affects transcription from the replication primer promoter. *Proc Natl Acad Sci U S A* **79**, 6313-6317 (1982).
26 Tomizawa, J. & Som, T. Control of ColE1 plasmid replication: enhancement of binding of RNA I to the primer transcript by the Rom protein. *Cell* **38**, 871-878 (1984).
27 Cesareni, G. & Banner, D. W. Regulation of Plasmid Copy Number by Complementary Rnas. *Trends Biochem Sci* **10**, 303-306, doi:Doi 10.1016/0968-0004(85)90168-9 (1985).
28 Castagnoli, L. *et al.* Genetic and structural analysis of the ColE1 Rop (Rom) protein. *The EMBO journal* **8**, 621-629 (1989).
29 Banner, D. W., Kokkinidis, M. & Tsernoglou, D. Structure of the ColE1 rop protein at 1.7 A resolution. *Journal of molecular biology* **196**, 657-675 (1987).
30 Eberle, W., Pastore, A., Sander, C. & Rosch, P. The Structure of Cole1 Rop in Solution. *Biol Chem H-S* **372**, 648-648 (1991).
31 Cohen, C. & Parry, D. A. D. Alpha-Helical Coiled Coils - a Widespread Motif in Proteins. *Trends Biochem Sci* **11**, 245-248, doi:Doi 10.1016/0968-0004(86)90186-6 (1986).
32 Munson, M., O'Brien, R., Sturtevant, J. M. & Regan, L. Redesigning the hydrophobic core of a four-helix-bundle protein. *Protein Sci* **3**, 2015-2022, doi:10.1002/pro.5560031114 (1994).
33 Munson, M. *et al.* What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. *Protein Sci* **5**, 1584-1593 (1996).

34 Castagnoli, L., Vetriani, C. & Cesareni, G. Linking an easily detectable phenotype to the folding of a common structural motif. Selection of rare turn mutations that prevent the folding of Rop. *Journal of molecular biology* **237**, 378-387, doi:10.1006/jmbi.1994.1241 (1994).
35 Cesareni, G., Cornelissen, M., Lacatena, R. M. & Castagnoli, L. Control of pMB1 replication: inhibition of primer formation by Rop requires RNA1. *The EMBO journal* **3**, 1365-1369 (1984).
36 Gregorian, R. S., Jr. & Crothers, D. M. Determinants of RNA hairpin loop-loop complex stability. *Journal of molecular biology* **248**, 968-984 (1995).
37 Eguchi, Y. & Tomizawa, J. Complex formed by complementary RNA stem-loops and its stabilization by a protein: function of CoIE1 Rom protein. *Cell* **60**, 199-209 (1990).
38 Eguchi, Y. & Tomizawa, J. Complexes formed by complementary RNA stem-loops. Their formations, structures and interaction with ColE1 Rom protein. *Journal of molecular biology* **220**, 831-842 (1991).
39 Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature biotechnology* **14**, 315-319, doi:10.1038/nbt0396-315 (1996).
40 Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **29**, 7133-7155, doi:Doi 10.1021/Bi00483a001 (1990).
41 Kellis, J. T., Jr., Nyberg, K. & Fersht, A. R. Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **28**, 4914-4922 (1989).
42 Woolfson, D. N. Core-directed protein design. *Curr Opin Struc Biol* **11**, 464-471, doi:Doi 10.1016/S0959-440x(00)00234-7 (2001).
43 Makhatadze, G. I. & Privalov, P. L. Energetics of protein structure. *Advances in protein chemistry* **47**, 307-425 (1995).
44 Willis, M. A., Bishop, B., Regan, L. & Brunger, A. T. Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure* **8**, 1319-1328, doi:Doi 10.1016/S0969-2126(00)00544-X (2000).
45 Chothia, C. Structural invariants in protein folding. *Nature* **254**, 304-308 (1975).
46 Chothia, C. The nature of the accessible and buried surfaces in proteins. *Journal of molecular biology* **105**, 1-12 (1976).
47 Bryson, J. W. *et al.* Protein design: a hierarchic approach. *Science* **270**, 935-941 (1995).
48 Hari, S. B., Byeon, C., Lavinder, J. J. & Magliery, T. J. Cysteine-free Rop: a four-helix bundle core mutant has wild-type stability and structure but dramatically different unfolding kinetics. *Protein Sci* **19**, 670-679, doi:10.1002/pro.342 (2010).
49 Clarke, J. & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**, 4322-4329 (1993).
50 Privalov, P. L. & Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of molecular biology* **86**, 665-684 (1974).

51 Bachhawat, K., Kapoor, M., Dam, T. K. & Surolia, A. The reversible two-state unfolding of a monocot mannose-binding lectin from garlic bulbs reveals the dominant role of the dimeric interface in its stabilization. *Biochemistry* **40**, 7291-7300 (2001).
52 Schellman, J. A. Selective binding and solvent denaturation. *Biopolymers* **26**, 549-559, doi:10.1002/bip.360260408 (1987).

Appendix A

Active Variants		R55F56	R55K	R55H	R55L	F55V56	W55I56
In vivo Screening		Active	Active	Active	Active	Active	Active
Thermal Melt	Tm (C)	66.81	63.75	59.95	70.5	62.37	66.23
Urea Melt	Cm (M)	4.3	3.8	1.8	4.4	2.403	2.101
Gibbs-Helmholtz Analysis	dH(kcal/mol)	103.37	106.10	103.80	107.70	103.61	93.23
	dCp(kcal/mol/K)	1.27	1.30	1.60	1.10	1.24	1.15
	Tg(K)	369.46	366.40	354.70	373.30	363.91	373.58
	dS(kcal/mol/K)	279.77	289.60	292.60	288.70	284.71	249.56
	Ts(K)	296.28	293.50	295.40	289.40	289.49	300.47
	dGs(kcal/mol)	10.61	11.00	8.90	12.60	11.00	9.45

Table A1: Thermal and Chemical Stability of In Vivo Active Variants and Resulting Gibbs-Helmholtz Analysis. * Represents data points still being collected

Intermediate Variants		R55Q	F55F56	W55Q56	L55S56	I55T56	R55A56	R55V56
In vivo Screening		Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate	Intermediate
Thermal Melt	Tm (C)	56.68	66.73	57.78	56.29	57.78	43.96	51.50
Urea Melt	Cm (M)	3.90	2.77	1.30	1.80	1.65	1.00	1.30
Gibbs-Helmholtz Analysis	dH(kcal/mol)	115.10	94.34	84.10	96.39	90.89	87.50	99.62
	dCp(kcal/mol/K)	1.50	1.14	1.06	1.28	1.13	1.26	1.45
	Tg(K)	354.00	375.54	366.14	363.32	367.19	352.76	355.63
	dS(kcal/mol/K)	325.20	251.20	229.71	265.29	247.54	248.04	280.11
	Ts(K)	284.10	301.41	295.05	295.27	294.73	289.95	293.30
	dGs(kcal/mol)	11.80	9.65	8.46	9.34	9.30	8.04	9.01

Table A2: Thermal and Chemical Stability of In Vivo Intermediate Variants and Resulting Gibbs-Helmholtz Analysis. * Represents data points still being collected

Inactive Variants		R55D56	R55R56	F55R56
In vivo Screening		Inactive	Inactive	Inactive
Thermal Melt	Tm (C)	36.56	30.98	43.02
Urea Melt	Cm (M)	0.80	0.70	1.20
Gibbs-Helmholtz Analysis	dH(kcal/mol)	74.32	60.83	74.85
	dCp(kcal/mol/K)	0.91	0.67	0.92
	Tg(K)	367.63	366.43	362.47
	dS(kcal/mol/K)	202.16	166.01	206.51
	Ts(K)	294.36	285.69	289.42
	dGs(kcal/mol)	7.68	6.98	7.83

Table A3: Thermal and Chemical Stability of In Vivo Inactive Variants and Resulting Gibbs-Helmholtz Analysis. The R55P variant was unfolded and CD signal showed non-helical structure. * Represents data points still being collected.

Appendix B

The Figures and Tables in Appendix B are the sole work of Anusha Kumar, the results were included in this research thesis because they help explain the interactions and effects of mutations at the R55F56 positions.

(a)

R55 Variants	T_m °C	$D^{1/2}$ (M)
L	70.5	4.4
I	66.4	3.7
R	66.8	4.3
K	63.8	3.8
M	66.8	3.5
W	64.8	-
A	63.5	3.8
V	61.5	-
H	60.0	-
Q	56.7	3.9
S	55.3	-
N	52.2	2.0
G	51.3	-
E	52.9	2.7
D	44.0	1.5

(b)

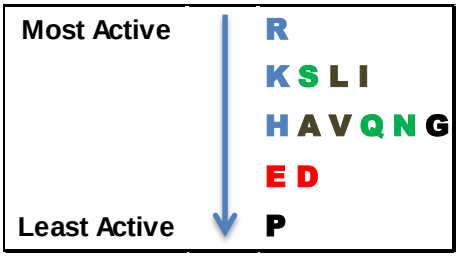


Figure B1: R55 Library Variants Activity, Tm, and D1/2. (a) T_m, melting temperature, $D^{1/2}$, is the chemical melting temperature with Urea. The color code is as follows: hydrophobic (brown), polar (green), positively charged (blue), negatively charged (red), special case (black).

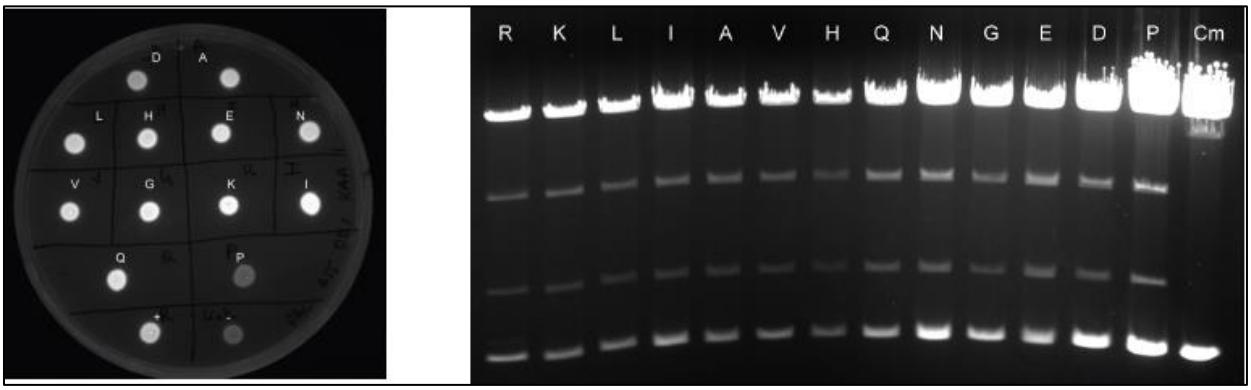


Figure B2: Screening results from the R55 Library. Figure on the (left) shows the fluorescence of the GFP as a direct indictor of Rop function. The figure on the (right) shows the an analytical digest of the pUCBADGFPuv and pAClacRop plasmids, the bottom and top bands are the pUCBADGFPuv plasmids with a ColE1 origin and the copy number is regulated by active Rop variants.

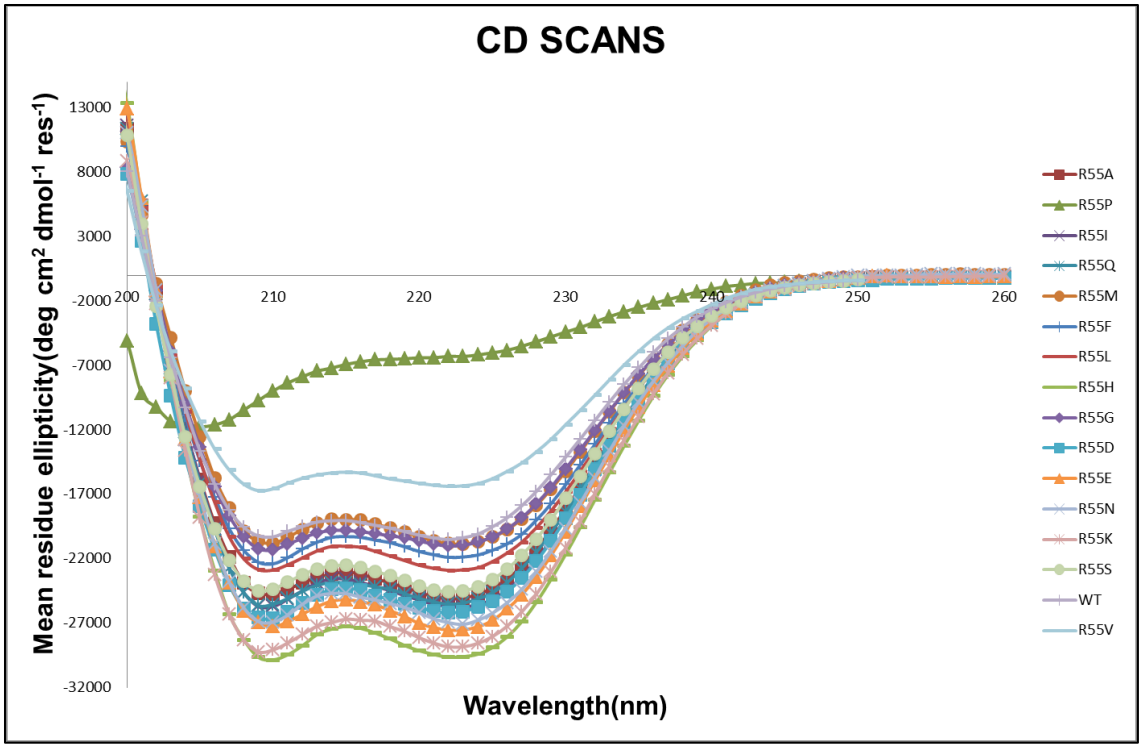


Figure B3: R55 Library Variants CD Scan of Mean Residue Ellipticity (MRE)

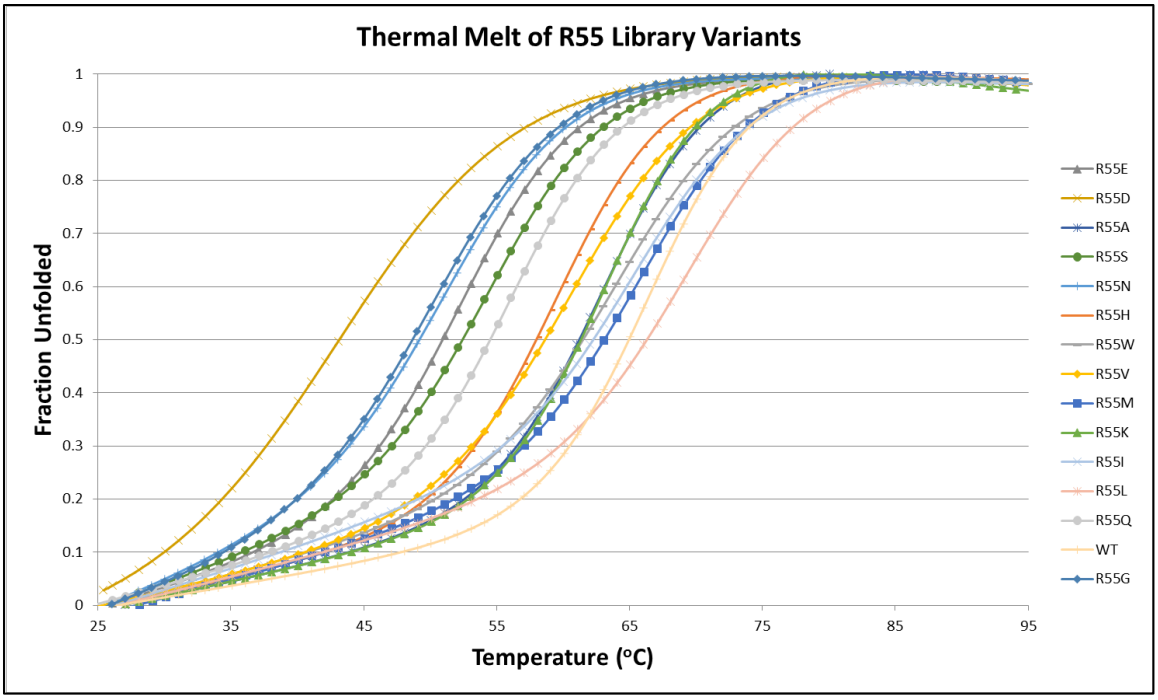


Figure B4: Thermal Denaturation of R55 Library Variants

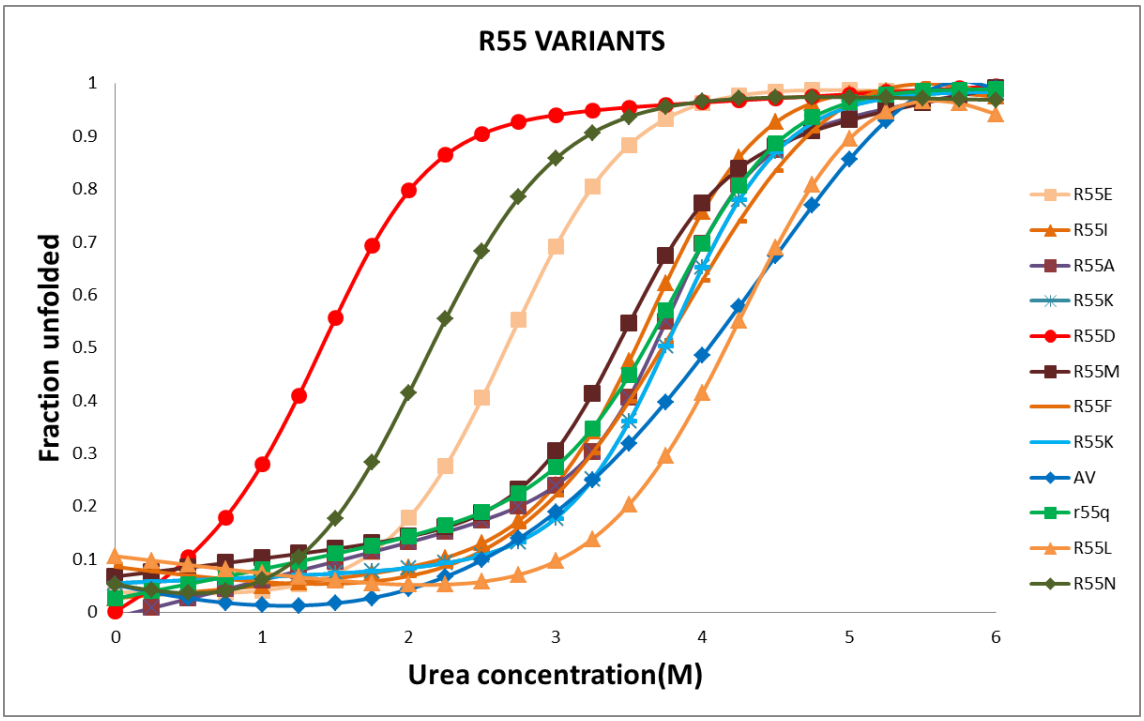


Figure B5: Chemical Denaturation by Urea of R55 Library Variants

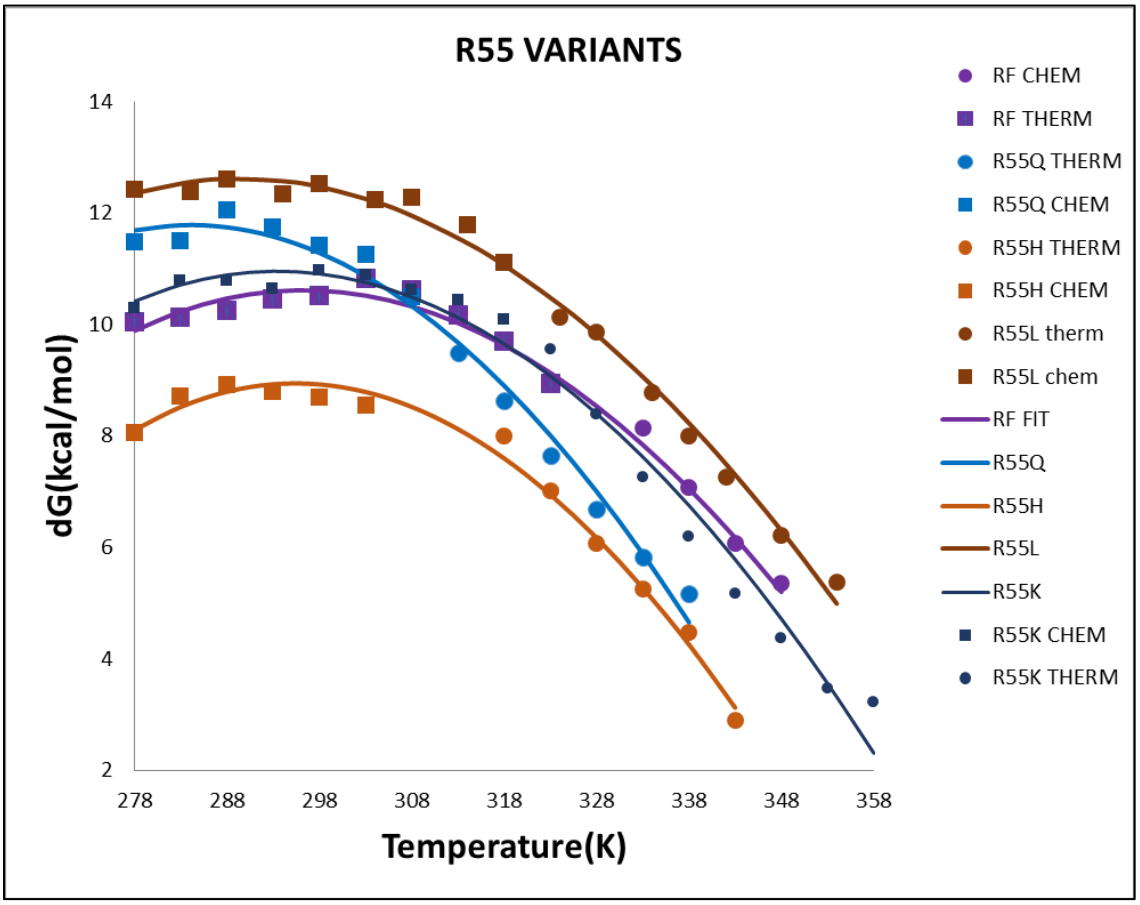


Figure B6: Gibbs-Helmholtz Analysis of the R55 Library Variants

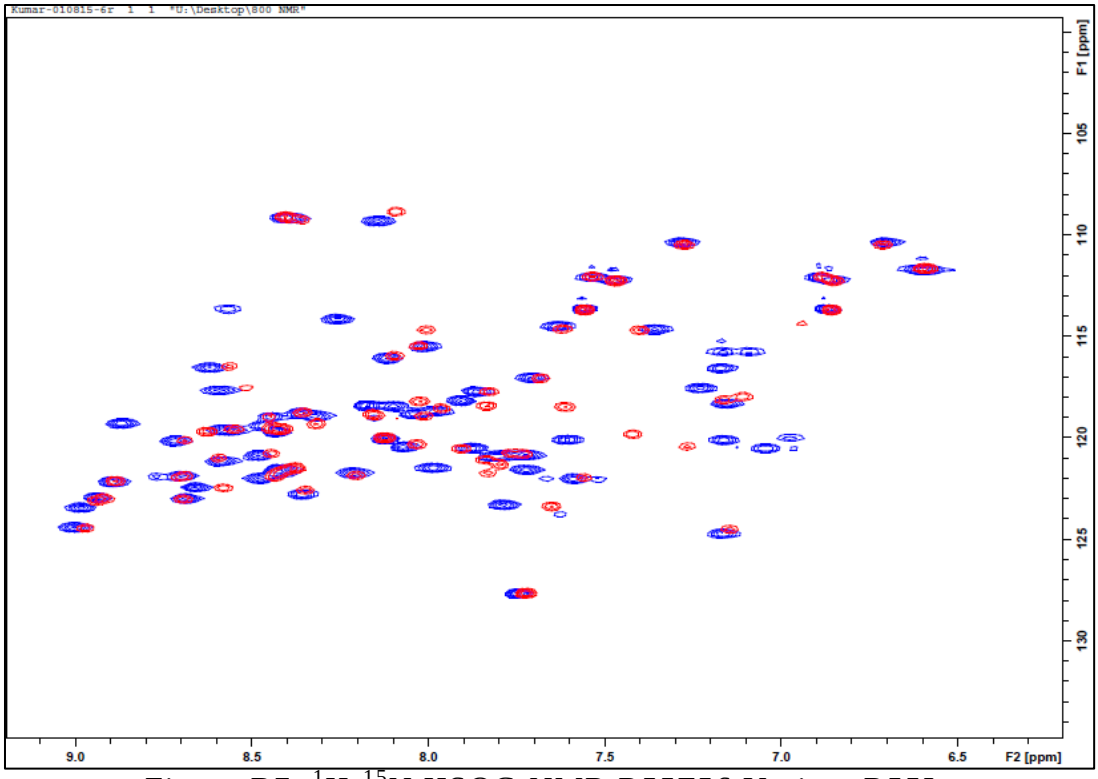


Figure B7: ^{1}H-^{15}N HSQC-NMR R55F56 Variant R55L.

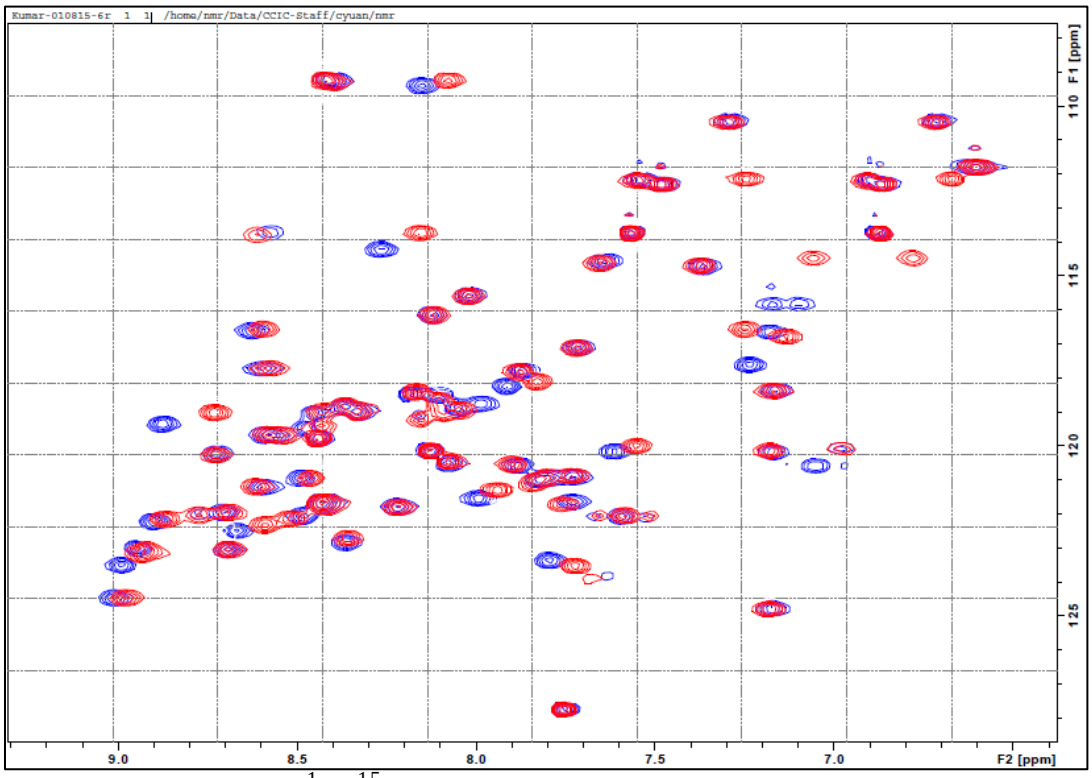


Figure B8: ^{1}H-^{15}N HSQC-NMR R55F56 Variant R55Q.

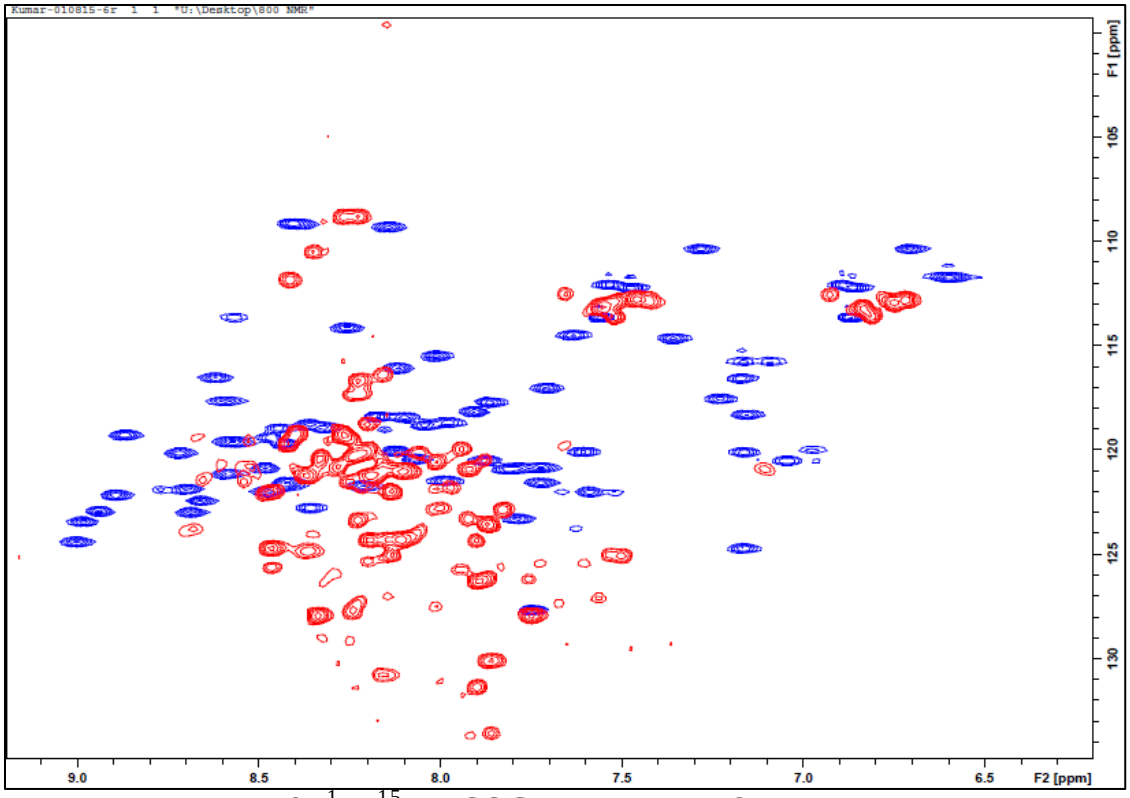


Figure B9: ^{1}H-^{15}N HSQC-NMR R55F56 Variant R55D.